March 19, 2020

Submitted via: www.regulations.gov

U.S. Environmental Protection Agency
EPA Docket Center
Mail Code 28221T
1200 Pennsylvania Avenue NW
Washington, DC 20460

Re: EPA Proposed Amendments to "National Emission Standards for Hazardous Air Pollutants: Miscellaneous Organic Chemical Manufacturing Source Category" Docket ID No. EPA-HQ-OAR-2018-0746 (84 Fed. Reg. 242; Dec 17, 2019)

To Whom It May Concern:

The Ethylene Oxide Panel of the American Chemistry Council (EO Panel), hereby submits comments on the proposed amendments to the "National Emission Standards for Hazardous Air Pollutants: Miscellaneous Organic Chemical Manufacturing Source Category" (MON). Our comments focus on amendments pursuant to the risk review that specifically address ethylene oxide (EO) including the following:

- MON Section IV. C. 3. **Determination of Risk Acceptability** (proposed MON amendment)
- Memorandum referenced in Section IV.C.3: **Sensitivity of Ethylene Oxide Risk Estimates to Dose-Response Model Selection.** 18 October 2019 from Paul White to Kristina A Thayer (ORD, 2019)

The MON (2019) proposal requests additional comments on the use of the 2016 Integrated Risk Information System (IRIS) unit risk estimate (URE) for ethylene oxide for regulatory purposes beyond those already submitted for the HCl Production RTR proposed rule as well as comments on the use of an alternative URE for ethylene oxide in the final rule for this source category. For reference, we attach the EO Panel's previous submissions, including its Information Quality Act Petition (2018), comments on the proposed HCL production RTR (HCL RTR, 2019), and comments on the Texas Commission on Environmental Quality's draft Decision Support Document for EO (TCEQ, 2019).

ED_005146_00001470-00001

The draft MON refers to the alternative URE proposed by TCEQ which was issued in June 2019. A revised draft and response to comments were released February 20, 2020 (TCEQ, 2020a,b), after submission of our initial set of comments. We strongly support the scientific approach used by TCEQ (2020a) to derive an alternative URE because it emphasizes biological plausibility and mode of action as guiding principles. The TCEQ (2020a) approach is based on the same National Institute for Occupational Safety and Health (NIOSH) cohort and lag period selected by IRIS (2016), and includes the same IRIS (2016) age-dependent adjustment factor (ADAF). Although this TCEQ (2020) alternative approach is different from that previously submitted by the EO Panel (2019) in comments on the proposed HCl RTR, they have in common the use of the standard log-linear Cox proportional hazard (CPH) model. The EO Panel (2019) approach is based on a combination of the NIOSH cohort and the Union Carbide cohort studies, utilizing all the available data from the two strongest cohorts.

This submission proposes two alternative URE's that are based on lymphoid mortality from the NIOSH cohort, including the TCEQ (2020a) proposed URE and a modification of the URE previously proposed by the EO Panel (2019) that now include the UREs from the NIOSH cohort alone (Table 1). This submission provides new figures and analysis of the observed epidemiology data to illustrate that the CPH model is more consistent with the dose-response form of the epidemiological, toxicological and biological mode of action. Our comments also include new scientific rationale for selecting lymphoid mortality[1], and not breast cancer incidence, as the critical endpoint for quantitative cancer risk. Although breast cancer outcomes should be considered as part of the weight of evidence for cancer assessment, breast cancer incidence in the full cohort or subcohort should not be used for quantitative cancer risk assessment because of the high potential for bias in the lower exposure range due to underascertainment of cases in the full cohort, most likely among workers who have shorter employment period and are harder to find (Steenland et al. 2003).[2]

[1] Steenland et al. (2004) published lymphoid mortality data, which IRIS (2016) converted to incidence data.

[2] Although Steenland et al. (2003) considers the subcohort of those interviewed to have complete ascertainment of breast cancer diagnoses, they were a subset of the full cohort, which was under-ascertained. There is no way of knowing that the distribution of cases by level of exposure in the interviewed population is comparable to the distribution in the fully ascertained total cohort. Steenland et al. (2003) indicated that there is greater difficulty of locating women with short term employment. Taken together, these data suggest a high potential for more cases missing at lower cumulative exposures. See section VI for further detailed discussion.

Table 1. Comparison of IRIS (2016) and New ACC Proposed Alternative UREs

	IRIS (2016) supralinear	ACC Alternative 1 Similar to TCEQ, 2020[1] Linear at 1/100[2]	ACC Alternative 2 Modified ACC,2019 HCl RTR Linear at 1/100[2]	Alternative 3 ACC,2019 HCl RTR Linear at 1/100[2]
Cohort	NIOSH Human	NIOSH Human	NIOSH and UCC Human	NIOSH and UCC Human
Critical endpoint	lymphoid incidence (based on mortality data[3]) in males and females and breast cancer incidence in females	lymphoid mortality[3] in males and females and males alone	lymphoid mortality[3] in NIOSH males alone is lowest central estimate of all cancer outcomes	lymphoid mortality[3] males and females
IRIS full model name	2-piece linear spline	Standard log-linear CPH	Standard log-linear CPH	Standard log-linear CPH
p-value	0.14 (includes all 3 parameters modeled)[4]	0.3	0.07	0.4
Lag Period	15 yr	15 yr	0 lag	0 lag
Age (yrs) limit[5]	85	70[5]	70[5]	70[5]
ADAF method[6]	IRIS (2016) approach apply 1.5 factor to slope[6]	IRIS (2016) approach apply 1.66 factor to slope[6]	EPA (2005b) cancer guidelines approach for each age[6]	EPA (2005b) cancer guidelines approach for each age[6]
Point of departure (PoD)	LEC 1/100	LEC 1/100,000	LEC 1/1,000,000	LEC 1/1,000,000

Rationale for PoD	Poor (none)[7]	Strong[7]	Medium[7]	Medium[7]
Model prediction of observed data	Poor[8]	Excellent[8]	Excellent[8]	Excellent[8]
URE (per ppm)	9.1	4.1E-03	3.3E-03	2.0E-03
1/M RSC (PPT)	**0.1**	**240**	**300**	**500**

Table 1 Footnotes

[1]**On** Feb 20, 2020, TCEQ (2020a) released a revised draft with a new proposed risk value similar to the ACC EO Alternative 1 proposed in ACC EO Panel comments submitted on Feb 19, 2020. To simplify, we use TCEQ's new value. A new alternative 2 includes NIOSH-only cancer values, rather than only values from combined NIOSH and UCC cohorts.

[2]**The** log-linear CPH model of excess relative risk (ERR), which has the general form ERR = exp(β C) – 1 in relation to exposure concentration C, is described as a "sublinear" model. In fact, this model becomes linear with slope = β as concentration C approaches zero, and it predicts ERR to be no greater than approximately 1% of that predicted by the corresponding linear model ERR= 1 + β C at all ERR levels less than approximately 0.02 (i.e., at ERR values ≤ 1/50).

[3]**Steenland** et al. (2004) only published lymphoid mortality data. IRIS (2016) converted this to lymphoid incidence by incorrectly applying the upper bound on the slope for cancer mortality to background incidence rates in a life-table calculation of the excess risk (Sielken and Valdez-Flores (2009a). Breast cancer incidence is not an appropriate critical endpoint based on strong potential for bias at lower cumulative exposures because a substantial number of cases could not be located, neither EPA nor the public has access to this data, and weaker findings compared to lymphoid cancer.

[4] **IRIS** (2016) p-value of 0.07 was corrected to 0.14. IRIS (2016) did not include the knot (point of inflection between the two linear regressions) as an estimated parameter in the 2-piece linear spline model. Also, IRIS (2016) did not include the lag as a parameter, so the p-values for Alternative values also does not include the lag as a parameter for more correct comparison. See text for detailed comments.

[5]**Age** limit for life table analysis. IRIS used 85 yrs, whereas TCEQ and ACC used 70 yrs consistent with default life span in EPA (2005a,b) cancer guidelines. See section VIII for detailed discussion.

[6]**IRIS** (2016) applied a single composite ADAF factor to the slope factor contrary to EPA (2006b) cancer guidelines which specifies that adjustments should be applied on an age-specific basis. TCEQ (2019, 2020) and ACC HCl RTR alternative risk values applied the ADAF factor correctly. However, consistent with EPA guidelines, the ADAF factor increased the risk on the year of exposure, but not for subsequent years, and resulted in an overall risk increase smaller than what would be expected in a model based on a constant exposure metric. TCEQ (2020a) decided to apply an ADAF factor similar to the IRIS (2016) approach. The composite ADAF factor for TCEQ is different from IRIS (2016) because TCEQ calculated risk at 70 yrs, while IRIS (2016) used 85 yrs. The composite ADAF factor is 1.66 for risks calculated at 70 yrs based on equation ADAF (70 yrs) =[(10 x 2 yr) + (3 x 14 yr) +54 yr]/ 70 yr = 1.66. ORD (2019) reports the ADAF factor to be approximately 1.5, which can be derived from equation ADAF (85 yrs)= [(10 x 2 yr) + (3 x 14 yr) +69 yr]/ 85 yr = 1.5

[7]**EPA** (2005a) cancer guidelines state that the PoD should be at the low-end of the observable range of exposures. IRIS did not justify the PoD of 1/100. TCEQ (2020, Appendix 7) proved 1/100 PoD is above the observable range for the log-linear mode, and 1/100,000 PoD is in the low-end of the observable range of exposures. 1/1,000,000 PoD was determined to be within the exposure range of individuals in the NIOSH cohort.

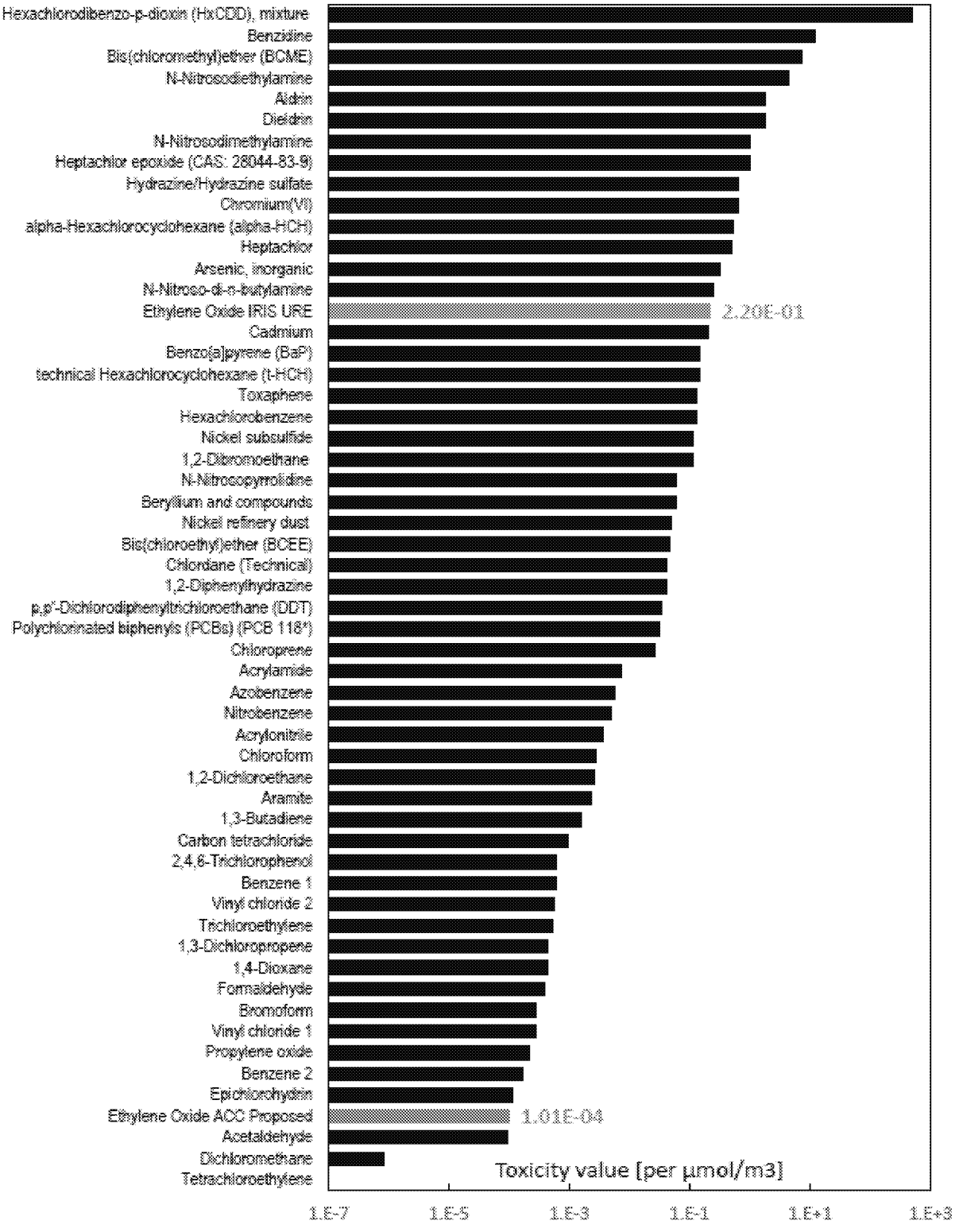
[8]**The** log-linear model but not the 2-piece spline models very accurately predict lymphoid cancer mortalities in the NIOSH data, whereas the confidence interval for the 2-piece linear spline significantly overestimates lymphoid cancer mortalities in the NIOSH data (TCEQ, 2020 Appendix 2).

In 2016, the EPA IRIS selection of the supralinear 2-piece spline model gave rise to one of the highest cancer potency estimates among those previously derived by IRIS (Figure 1). The EO Panel presents evidence that the very steep slope of this model is not justified based on the relatively weak epidemiological findings reported by Steenland et al. (2003, 2004) and the weight-of-evidence in the epidemiological literature (Marsh et al. 2019). The IRIS (2016) URE for EO results in a 1 in a million extra risk specific concentration (1/M RSC) for EO of 0.1 ppt, which is highly implausible based on epidemiological, toxicological and biological mode-of-action evidence. IRIS (2016) considered an unprecedented number of statistical models each with different permutations of lag times, exposure metrics and, in the case of the spline models, multiple positions of the knot. The selection of the model was based almost exclusively on a fundamentally flawed statistical analysis and a flawed assessment of visual fit in relation to categorical data without consideration of biological plausibility.

The proposed MON relies on the ORD (2019) memo that includes a sensitivity analysis evaluating a range of alternative risk values, concluding that the IRIS upper-bound URE values could have been up to 5 times lower. We agree with consideration of a range of values including central estimates, but the ORD memo rejects a more standard statistical model—the log-linear CPH model—that has a comparable statistical and visual fit to the one selected by IRIS[3]. More importantly the CPH model is a simpler model that has a dose-response form with greater biological plausibility and is more consistent with the observed epidemiological data, fitting the EPA SAB (2015) selection criteria for models. This is the same model used to derive alternative values by TCEQ (2019) and ACC (2019) based on Valdez-Flores et al. (2010).

[3] TCEQ (2019 table 38) calculates a correct p-value of 0.14 for the IRIS-selected 2-piece spline model fit to lymphoid data. Thus, neither the IRIS-selected 2-piece spline model nor a corresponding fitted CPH model are statistically significant. Despite the lack of significance for an exposure-response relationship, a conservative yet scientifically sound alternative approach is to calculate extra risk using the CPH model. The CPH model becomes the model of choice because it is a more parsimonious (simpler) model, has greater biological plausibility, and better predicts the observed lymphoid mortalities in the NIOSH study compared to the IRIS selected 2-piece spline model.

Figure 1 The EO URE is among the highest IRIS inhalation UREs for known or likely carcinogens



The ORD (2019) sensitivity analysis includes a non-statistically significant linear regression fit to the categorical data, which the SAB (2015) specifically rejected because it was based on categorical results instead of the continuous individual-level exposure data. ORD (2019) also rejects a linear model of the individual data for lymphoid cancer while including this same linear model for breast cancer. Thus, the ORD (2019) memo "stacks the deck" by only including models that result in UREs within 5-fold of the IRIS (2016) URE, while excluding models that adhere to SAB's principles, but have combined cancer UREs 15- to 40-fold lower than the IRIS (2016) URE (Table 2). These comparisons are based on IRIS (2016) estimates and methods for applying the CPH and linear models using lymphoid and breast cancer incidence. If the EPA (2005) cancer guidelines is applied more correctly (discussed in detail in previous comments submitted on the proposed HCl RTR rule and summarized in Table 1 footnotes of these comments), then the IRIS URE is about 2000-fold lower than the IRIS value (Table 1).

A major flaw with IRIS (2016) and ORD (2019) approaches that led to rejection or marginalization of the continuous exposure linear and standard Cox-proportional log-linear models is that they relied on subjective visual fit based on misleading figures (IRIS Figures 4-3, 4-8) that compare the models with very few (4 to 10) categorical (grouped) rate[4] ratio data points that are not representative of the larger set of individual data modeled (53 lymphoid and 233 breast cancer cases). These figures are incorrect not only because a few grouped categorical rate ratios are used to represent the individual rate ratios, but also because they misrepresent the actual data (i.e. hazard rates) modeled along the y-axis.

Although IRIS (2016) added a note to the figure legends that the relative rates (RR) of the different models are not strictly comparable along the y-axis[5], this note has not been adequate to prevent misinterpretation of these figures leading to incorrect conclusions about model fit. The **ORD (2019) memo included similar misleading figures, but did not add the very brief but important warning, strongly suggesting that ORD (2019) may have made similar errors in rejecting the CPH model based on invalid model comparisons along the y-axis.** It is notable that the draft IRIS (2013) assessment reviewed by SAB (2015) also did not include this very important warning.

[4] Steenland et al. (2003, 2004) refer to these as odds ratios, but they are technically rate ratios. IRIS (2016) refers to them as odds ratios and relative rates (i.e. rate ratios), or more generally, relative risk.
[5] IRIS (2016) figures include the following parenthetical note: "Note that, with the exception of the categorical results and the linear regression of the categorical results, the different models have different implicitly estimated baseline risks; thus, they are not strictly comparable to each other in terms of RR values, i.e., along the y-axis."

Table 2. Comparison of IRIS (2016) derived models using ORD (2019, Table 1) approach (UREs do not include the ADAF)

	2-piece linear spline Supralinear	Linear regression (forcing to 1 at 0 exposures)	Linear	Standard CPH Linear at EPA POD 1/100
Model of individual data?	Yes	NO	Yes	Yes
IRIS full model name	Linear spline model with knot at 1,600 ppm x days	Linear regression of categorical results	Linear	Log-linear model (standard Cox regression model)
LYMPHOID INCIDENCE (Males and Females)				
IRIS p-value	0.14 corrected from 0.07	0.18	0.13	0.22
Central estimate URE (per ppm)	1.34	0.44	0.02	0.0095
Upper bound URE (per ppm)	5.26	0.97	0.083	0.020
1/1,000,000 RSC with ADAF of 1.5 (ppt)	**0.1**	**0.7**	**7.8**	**31.8**
BREAST CANCER INCIDENCE (Females)				
IRIS p-value	0.04 corrected from 0.01	0.16	0.01	0.02
Central estimate URE (per ppm)	0.71	0.42	0.19	0.08
Upper bound URE(per ppm)	1.48	0.91	0.38	0.14
These data are not appropriate for quantitative risk assessment purposes because authors report substantial number of missing cases with higher potential for those with shorter employment missing (Steenland et al. 2003). These data have not been available for independent evaluation by EPA or the public, and, thus, lack transparency, verification and independent analysis.				
LYMPHOID & BREAST CANCER INCIDENCE (Males and Females)				
Central estimate URE (per ppm)	2.1	0.9	0.2	0.1
Upper bound URE (per ppm)	6.1	1.6	0.4	0.15
1/1,000,000 RSC with ADAF of 1.5 (ppt)	**0.1**	**0.3**	**1.0**	**2.9**
These data are not appropriate for risk assessment because the breast cancer incidence data are included. EPA provided no justification for the POD of 1/100. TCEQ (2020a) analysis shows that the POD 1/100 for the standard CPH model extrapolates above or in the high range of the experimental data!				

A superior approach to assessing model fit is to calculate the number of cases predicted by the model statistically rather than to relying on subjective "eyeballing" of the data. TCEQ (2020a) used this more objective approach to show unequivocally that the CPH and linear models predict the cases observed as a whole and in the lower cumulative exposure levels compared to the IRIS- selected 2-piece spline model which consistently overestimate. Based on this analysis, the CPH and linear models have greater local and overall fit compared to the IRIS (2016) selected 2-piece linear spline model.

A second major flaw of the proposed MON (2019) and ORD (2019) memo is they are based on incorrect statistics of the IRIS (2016) report. The ORD (2019) memo points out (at p. 6) that "It is important to note that this analysis relies entirely on results and equations presented in the final EO IRIS assessment", and thereby makes clear that it did not independently evaluate IRIS (2016) statistical analysis or consider TCEQ's peer review of the IRIS analysis. The proposed MON rule notes the concerns raised by TCEQ but appear to dismiss them by claiming that the proposed TCEQ assessment has not been peer reviewed[6]. However, TCEQ points out a very simple statistical error: IRIS (2016) did not account for all three instead of just two parameters that were numerically optimized for all spline models[7]. The mistake of omitting a single degree of freedom led IRIS (2016) and ORD (2019) to mischaracterize the supralinear 2-piece spline model fits as being adequate while rejecting the CPH model.

The simple statistical error documented by TCEQ is not the type of claim that requires peer review to be valid. It is a basic principle clearly stated in the National Research Council report entitled "Models in Environmental Regulatory Decision Making", which states that the strategy to pick the "best model" for regulatory decision making should be "*subject to a penalty function reflecting the number of model parameters, thus effectively forcing a trade-off between improving model fit by adding addition[al estimated] model parameters versus having a parsimonious description*" (NRC, 2007, pp. 174). The EO Panel provides new information revealing that IRIS (2016) incorrectly claims the SAB approved this apparent violation of basic statistical principles.

---

[6] The MON (2019) states: "TCEQ highlighted uncertainties in the URE arising from what it considered to be errors in the assumptions and calculations used to determine the best model fit on the data. TCEQ's concerns with the EPA's URE derivation have not been peer reviewed and the public comment period closed on September 26, 2019." Fed Reg 2019 (Tues Dec 17); 84 (242):69218

[7] The SAS statistical software used by IRIS required the user to ensure that a correct number of estimated parameters is entered into (or assumed by) that program when fitting a model to data. In the case of 2-piece spline model fits, the parameters representing the two slopes and the X-axis value of the "knot" or point of intersection that connects those two slopes).

We bring to light an analysis by IRIS (2016, Appendix D) estimating p-values with and without including the parameter for the knot demonstrating that IRIS was aware of this issue. **The fact that a simple statistical error of omitting a single degree of freedom and errors in evaluating visual fit of models can result in a 10- to 260-fold difference in EO cancer potency for each cancer based on IRIS assumptions and calculations for the CPH model emphasizes the tenuous basis of any EPA rule, such as the MON, that relies on the IRIS (2016) EO cancer assessment (Table 2)[8].**

The MON also importantly notes the SAB advice that model selection should have a "dose-response shape that is… biologically plausible" (MON, footnote 39). In this context, as an important dose perspective, our bodies produce EO through normal metabolic processes at levels that are approximately equivalent to inhalation of 1,900 ppt ± 1,300 ppt (Kirman and Hays, 2017). The IRIS (2016) stated that "it is *highly plausible* that the dose-response relationship *over the endogenous range* is sublinear" [emphasis added]. The basis for this conclusion was described in our comments on the HCL RTR.

Briefly, EO molecular and tissue injury is moderated at low EO exposures by overlapping biological defenses. These new comments add additional information that demonstrate that none of these biological defenses are plausibly expected to be saturated at low EO exposures. These data further indicate that it is highly biologically implausible that the contribution of an additional 0.1 ppt exogenous EO to an existing 1,900 ppt background endogenous EO exposure, would result in a sudden and biologically unexplained shift to a supralinear exposure response and mode of action. This is particularly so considering that such an additional minute exogenous EO exposure is also a very small fraction of the reasonable variability range of normal human endogenous background EO exposures (1,300 ppt). We provide further evidence supporting the conclusions of Kirman and Hays (2017) that the standard deviation reflects expected biological variation and not experimental variation.

The biological plausibility of a low-dose supralinear dose response is also inconsistent with animal toxicology and mode of action data for EO. Of particular importance is that ethylene is not carcinogenic in rats despite producing an approximate 3 ppm EO equivalent dose at the top 3,000-ppm-ethylene tested exposure. The IRIS supra-linear EO dose response incorrectly predicts that the ethylene bioassay should have been positive. The biological plausibility of the EPA hypothesized low-dose supra-linear dose response is inconsistent with the observation that doses of EO in rats did not increase DNA adducts, the molecular-initiating mode of action target of EO-induced cancer, at approximately 4 orders of magnitude greater than the dose in a human exposed to 0.1 ppt EO (Marsden et al., 2009).

[8] See footnote 1 on page 3 for further details

In summary, the proposed MON relies on the ORD (2019) sensitivity analysis which is based on the IRIS (2016) incorrect statistical analysis and misleading visual fit comparisons. When correct statistics, parsimony (i.e. priority for simpler models), biological plausibility, consistency with the observed data and the weight of evidence from the epidemiology studies are considered fully, the standard log-linear CPH model is the most appropriate model that adheres closely to SAB (2015) principles for model selection. The following detailed comments elaborate on these points:

I. **The proposed MON is based on the ORD (2019) visual fit comparisons to categorical data, which misrepresents the individual data modeled. This flawed visual fit as the basis for the IRIS (2016) selection of a spline model leads to deriving one of the highest inhalation UREs.**

II. **The ORD (2019) memo did not make a simple correction in statistical analysis that led to an incorrect conclusion that the 2-piece spline model has superior fit compared to the CPH model. Because the ORD memo—cited as a key basis for the proposed MON rule—is flawed, so is the proposed MON rule determination of risk acceptability and uncertainty.**

III. **ORD (2019) lists SAB (2015) recommendations as the basis for selection of the 2-piece model and alternative models. However, ORD (2019) inclusion of linear regression of categorical data and exclusion of the linear and log-linear CPH models are internally inconsistent and contradictory to SAB (2015) recommendations. The proposed MON rule should include these alternative values, which would result in 15- to 40-fold lower risk values for combined lymphoid and breast cancer incidence, and 140- to 260-fold lower for lymphoid incidence alone. These estimates are based on IRIS (2016) assumptions and methods for deriving upper-bound UREs, and not the EO panel's approach.**

IV. **The ORD (2019) sensitivity analysis does not consider biological plausibility and consistency based on the results of the epidemiological studies. The overall weak findings suggest a shallow and not a steep exposure response at low exposures, and do not support derivation of one of the highest inhalation URE's.**

V. **Lymphoid morality in humans is an appropriate health outcome for risk assessment, as is, without transformation to incidence. The overall weak findings of the lymphoid mortality data suggest a shallow and not a steep exposure response at low exposures. These data are not consistent with the IRIS derivation of one of the highest inhalation UREs.**

VI. Although useful for consideration of the overall weight-of-evidence, breast cancer should not be considered a critical cancer endpoint based on the lack of robust findings in Steenland et al. (2003, 2004) and weight-of-evidence in the epidemiological literature. Breast cancer incidence data should not be used for quantitative risk assessment purposes because of substantial under-ascertainment of cases reported by Steenland et al (2003).

VII. IRIS (2016) did not consider the biological plausibility of models based on biological mode of action and toxicological evidence, which support a shallow linear exposure-response at lower exposures. IRIS has not offered any biologically plausible mode of action analysis accounting for a supralinear dose-response of EO in the low-exposure range. In contrast, considerable experimental mode of action data consistently indicate it is highly implausible that EO operates by supralinear exposure response in the exposure region estimated by IRIS as increasing cancer risks. The IRIS (2016) risk specific concentration of 0.1 ppt is overly conservative to the point of lacking regulatory utility because it is 4 orders of magnitude lower than average human background (predominately endogenous) exposure levels and variability.

VIII. The ACC alternative proposal for URE is conservative and has a dose-response form that is both biologically plausible and consistent with the observed data. The rationale for selecting a critical endpoint and point-of-departure are summarized.

EPA must revise its risk modeling and analysis by using a unit risk estimate (URE) value for EO that is scientifically justified instead of relying on the 2016 IRIS value. The IRIS value is an incorrect and overly conservative value based on implausible exposure-response models. We strongly encourage EPA to consider all of the available alternatives to the EO IRIS value for regulatory uses. Thank you.

Sincerely,

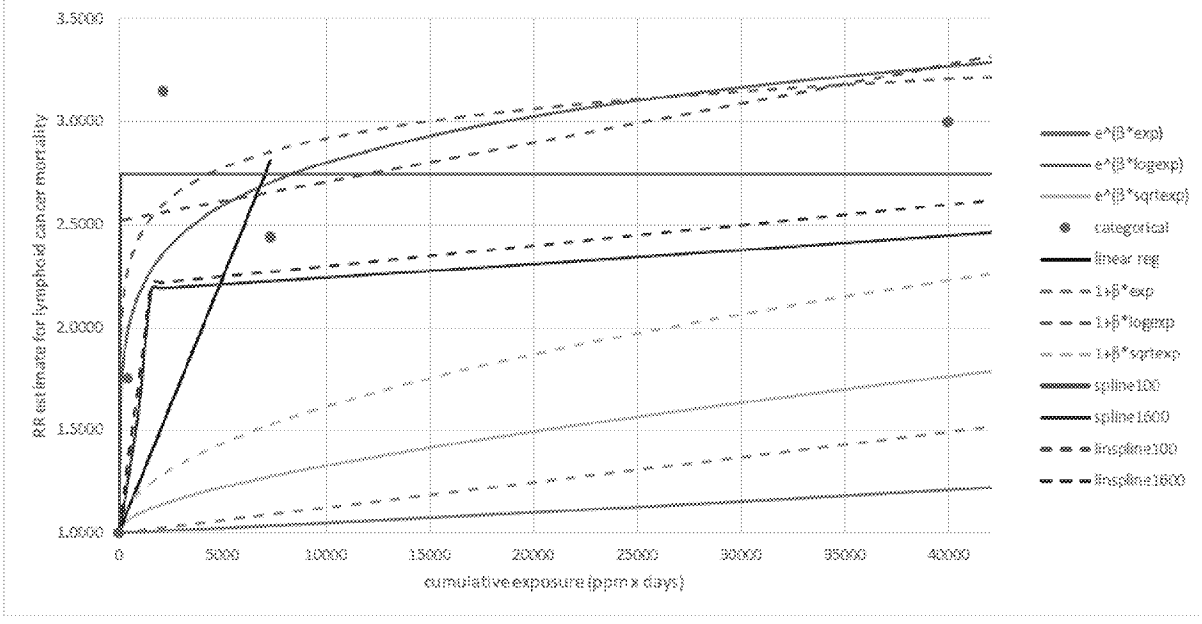
William Gulledge

William P. Gulledge
Senior Director
Chemical Products & Technology Division

I. The proposed MON is based on the ORD (2019) visual fit comparisons to categorical data, which misrepresents the individual data modeled. This flawed visual fit is the basis for the IRIS (2016) selection of a spline model leading to the derivation of one of the highest inhalation UREs.

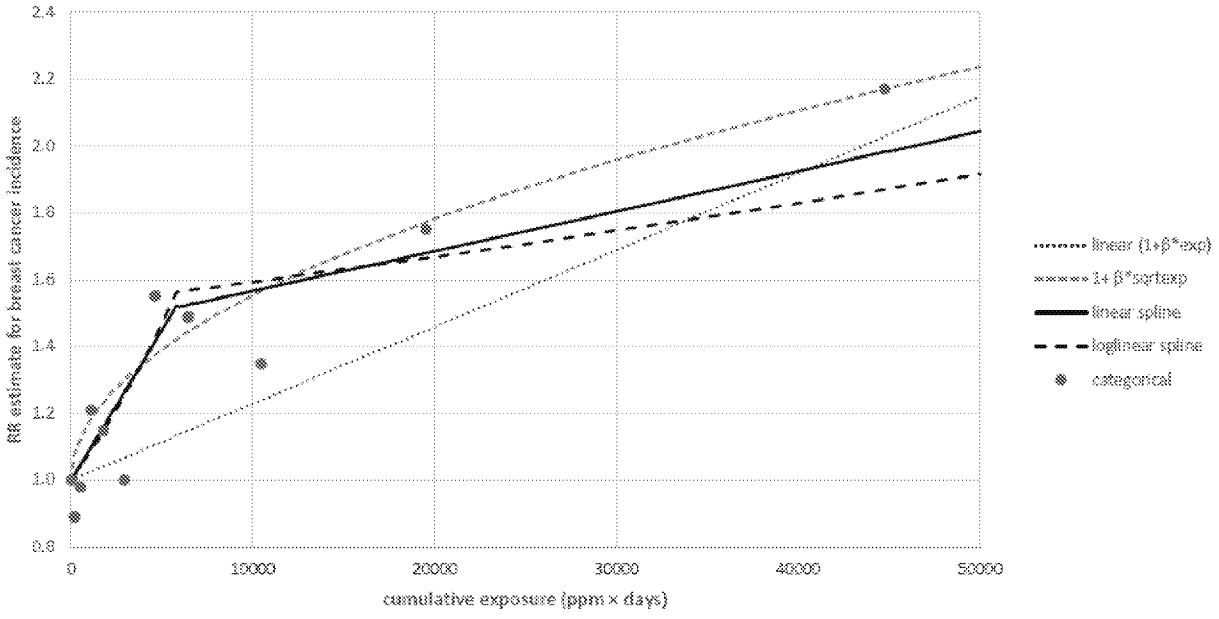
Steenland et al. (2003, 2004) calculated four to ten odds ratios (ORs; i.e. categorical rate ratios, RR), which uses worker to worker comparisons within the study. These odds ratios play a major role in the derivation of a very high URE because ORD (2019) and the IRIS (2016) both use very few odds ratios to determine which continuous models (e.g. 2-piece spline vs. CPH) have acceptable "local fit" at the lower exposure levels based on subjective visual comparisons. Figures 2 and 3 are identical to Figures 4-3 and 4-8 from the IRIS (2016), which use the grouped "categorical" odds ratio data (solid purple points) to compare the visual fits of the different models. These figures give the false impression of a very clear dose response pattern when confidence intervals (CI) are not added. CIs that overlap "1" are not statistically significant. The same categorical data with CIs are shown first on the same y-axis linear scale as the IRIS figures (Figures 4 and 5), and then on a log scale (Figures 6 and 7).

Figure 2 IRIS (2016) Figure 4-3: Odds ratio for quartiles (closed circles) for lymphoid cancer mortality (with 15 year lag)



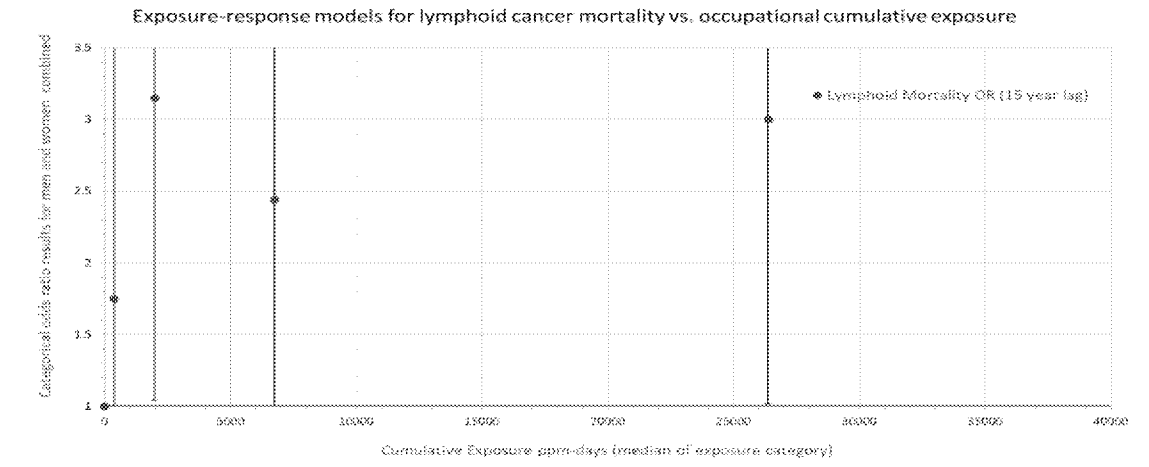
The red dashed line is the 2-piece spline model with knot at 1600 ppm-days selected by IRIS. The solid blue line is the CPH model proposed in this report and published by Steenland et al. 2004. Note: IRIS states that the various models are not comparable along the y-axis.

Figure 3 IRIS (2016) Figure 4-8: Odds ratio for deciles (closed circles) for breast cancer incidence in subcohort with interviews (with 15 year lag)



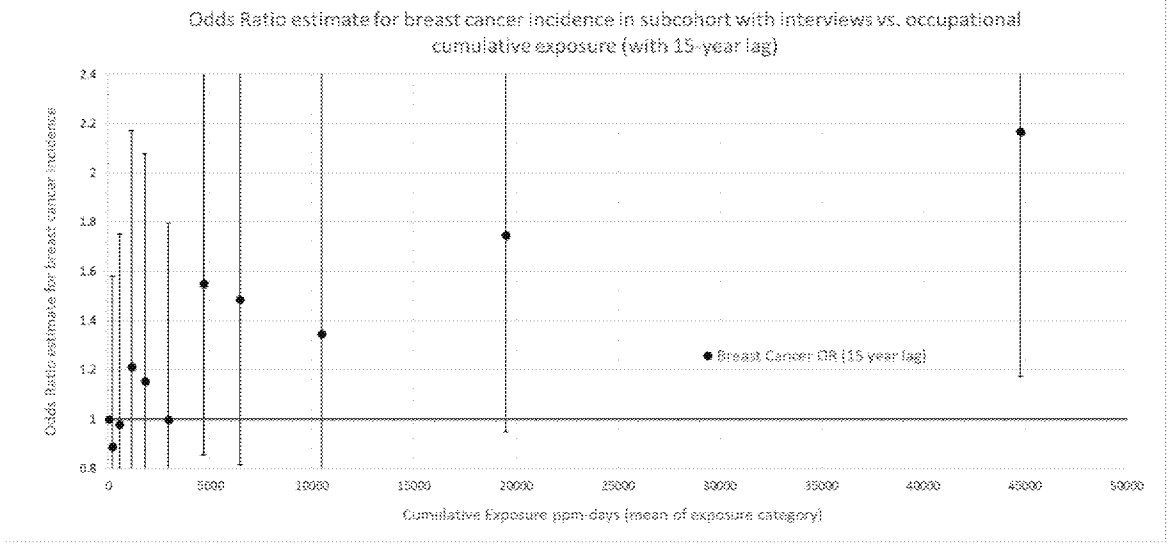
The black solid line is the 2-piece spline model selected by IRIS. IRIS did not plot the CPH model for this subcohort with interviews, but it would be flatter than the red dotted line. The purple dots are the categorical RR (i.e. odds ratio). Note: IRIS states that the various models are "not strictly comparable along the y-axis"

Figure 4 Odds Ratio for Lymphoid Mortality (15 year lag, males and females) with same y-axis linear scale as IRIS (2016) figures used to determine visual fit (Odds ratio and 95% CI)



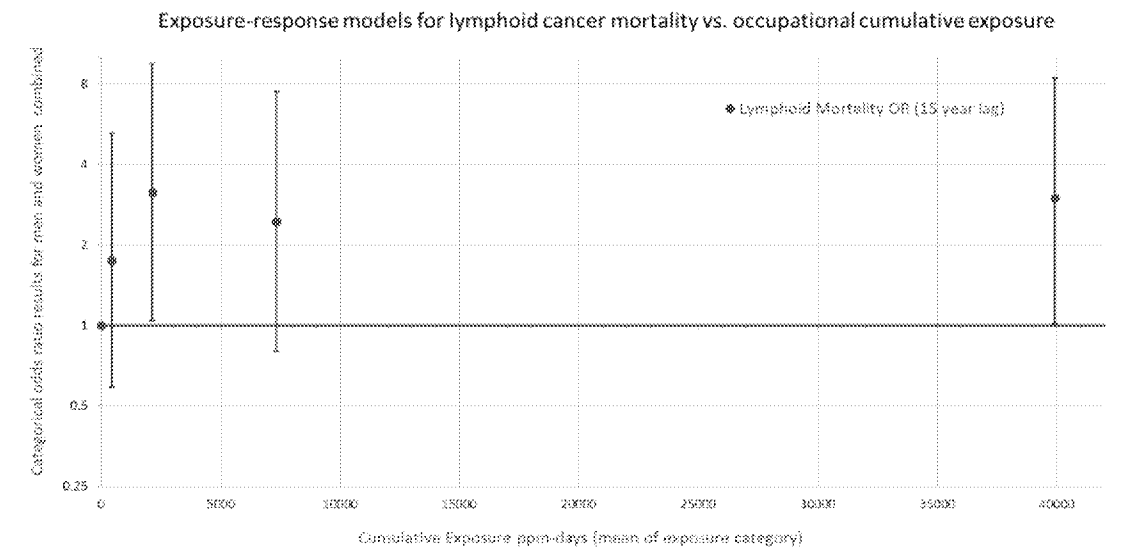
Data from IRIS (2016) Tables D-26 and D-28. ORs with confidence intervals indicate exposure response is not supralinear. Medians of exposure category were reported and are considered superior to mean values. ORs with CIs that do not include 1 are statistically significant.

Figure 5 Odds Ratio for Breast Cancer Incidence (15 year lag, females, subcohort) with same y-axis linear scale as IRIS (2016) figures used to determine visual fit (Odds ratio and 95% CI)



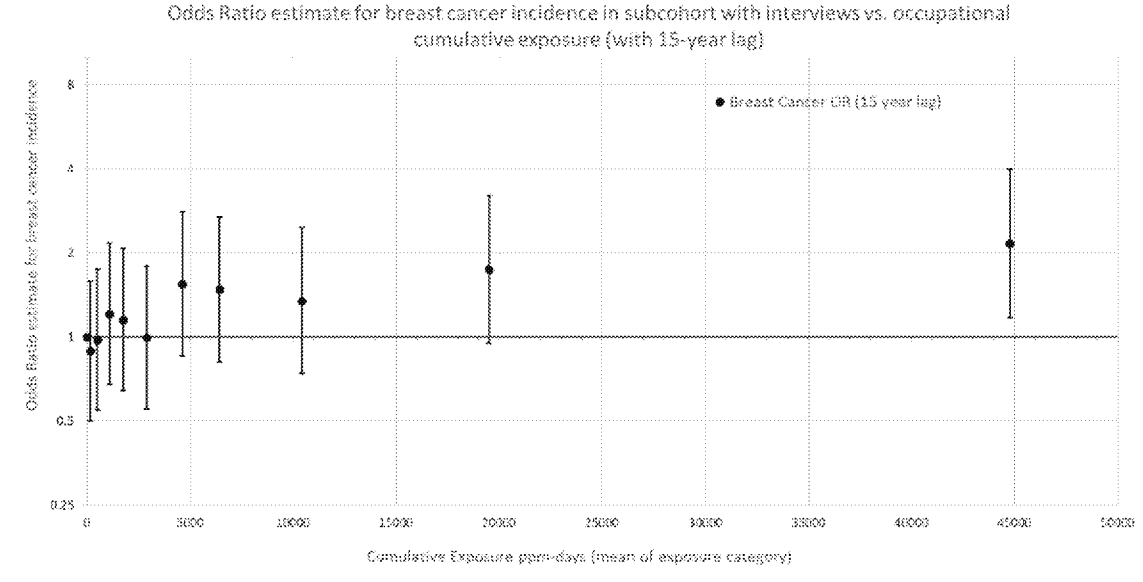
Data from IRIS (2016) Table D-1 and D-3. ORs with confidence intervals (CI's) are not consistent with a supralinear exposure response model. Medians of exposure category were not reported by IRIS. ORs with CIs that do not include 1 are statistically significant.

Figure 6 Odds Ratio for Lymphoid Cancer Mortality (15 year lag, both sexes) with log scale (Odds ratio and 95% CI).



Data from IRIS (2016) Tables D-26 and D-28. ORs with CIs are not consistent with a supralinear exposure response model. ORs with CIs that do not include 1 are statistically significant

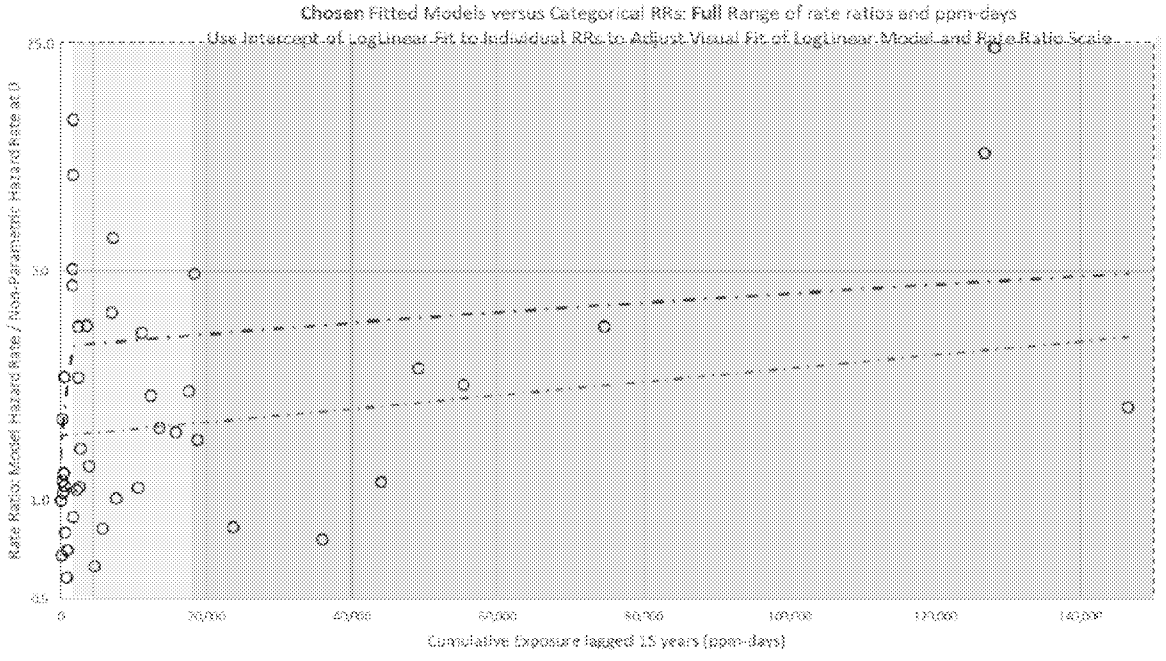
Figure 7 Odds Ratio for Breast Cancer Incidence (15 year lag, females, subcohort) with log scale (Odds ratio and 95% CI)



Data from IRIS (2016) Tables D-26 and D-28. ORs with CIs are not consistent with a supralinear exposure response model. ORs with CIs that do not include 1 are statistically significant

More importantly, the IRIS figures (Figures 2 and 3) plot data points that represent grouped "categorical" data aggregated into quartiles or deciles, instead of the actual individual cases (53 lymphoid and 233 breast cancer) modeled as shown below in Figure 8. Although Figure 8 will also have very wide confidence intervals, it is far more representative of the exposure response pattern of the individual data that are modeled compared with Figures 2 and 3 that use only the 4 and 10 data points used by IRIS to evaluate visual fit. The latter data plots with few categorical data points mask the true but more noisy exposure-response relationship for lymphoid mortality shown below in Figure 8 for lymphoid mortality and described in greater detail by Valdez-Flores and Sielken (2013) for breast cancer mortality.

Figure 8 Lymphoid mortality comparing representative odds ratios (open circles) for individual cases for the purpose of comparing visual fit with the CPH (blue line)and supralinear 2-piece spline model (red line) adjusted along the y-axis for a more correct comparison (TCEQ, 2020b Figure 6).



In this figure, the plotted fits are directly comparable, showing that both fits are roughly similarly consistent with the data. An apples-to-apples comparison is shown here of 2-piece spline and CPH modelled rate ratios adjusting the y-intercept for the differences in the estimated baseline risks. In contrast to IRIS (2016) Figure 4-3 (Figure 2 in these comment above), this corrected plot addresses the note in IRIS (2016) Figure 4-3 (and in similar figures used by IRIS to compare visual fits) warning that "the different models have different implicitly estimated baseline risks; thus they are not strictly comparable to each other in terms of RR values (i.e. along the y-axis)." The different shaded areas represent the variable exposure range for each category of exposure that IRIS (2016) used to achieve equivalent size groups of cases.

In addition to misrepresenting the modeled data with few categorical odds ratios, the graphs also are extremely misleading along the y-axis as stated in the figure legends in IRIS (2016) as follows:

> Note that, with the exception of the categorical results and the linear regression of the categorical results, the different models have different implicitly estimated baseline risks; thus, they are not strictly comparable to each other in terms of RR values, i.e., along the *y*-axis. They are, however, comparable in terms of general shape.

This cautionary note indicates that the visual model comparisons between spline vs. linear or CHP model fits to categorical data are invalid due to non-comparability of values plotted along the y-axis. ORD (2019) failed to include this extremely important statement in the figures used to evaluate local fit, suggesting ORD (2019) may have been misled by these figures along the y-axis. This statement was also not included in the IRIS (2013) draft assessment reviewed by SAB, and yet these figures appear to be an important basis for SAB's general agreement with IRIS (2013) proposed selection of the 2-piece linear spline model for breast cancer incidence based on "local fits in the low exposure range"[9].

Figure 8 from TCEQ (2020b) correctly adjusts the visual representation of the models for the difference between the implied background hazard rates of the models and the non-parametric background hazard rate. It more correctly and fairly compares the supralinear spline model (red dashed curve) and the standard CPH model (blue dashed curve). Figure 8 adjusts the graphs of the models along the y-axis for illustrative purposes to account appropriately for different estimated baseline hazards relied on by each of the plotted models. Unlike the impression given by Figure 2, Figure 8 shows that the CPH model fits to individual-level lymphoid data do not systematically underpredict those data compared to 2-piece spline fits, either specifically at low exposures or over the entire exposure range. All of these models reasonably characterize the general (fairly noisy) pattern of the data.

In summary, model selection should not be based on invalid visual comparisons of model fits to categorical data. When odds ratios with the CIs are considered, the original epidemiology study together with the weight of evidence reviewed in our comments on the proposed HCl RTR rule do not support a supralinear slope or the consequent derivation of one of the highest UREs published by IRIS for inhalation cancer risks, including comparisons to unequivocal human carcinogens (Figure 1).

[9] SAB (2015) generally concurred with 2-piece spline based on "local fit" for breast cancer incidence. No specific recommendation was given for lymphoid mortality. Good local fit to the low-exposure was based on the misleading figures and comparisons with categorical data without confidence intervals included.

II. The ORD (2019) memo did not make a simple correction in statistical analysis that led to an incorrect conclusion that the 2-piece spline model has superior fit compared to the CPH model. Because the ORD memo—cited as a key basis for the proposed MON rule—is flawed, so is the proposed MON rule determination of risk acceptability and uncertainty.

Based on the statistical analysis presented in IRIS (2016), ORD (2019) claims that the 2-piece spline model has superior fit compared to the CPH model. However, this conclusion is based on incorrect statistical fit calculations in the IRIS (2016) for the spline due to a simple error of not accounting for all the parameters. Prior to the ORD (2019) memo, TCEQ (2019) provided a detailed and complete evaluation of the statistical analysis presented in the IRIS (2016) appendices. TCEQ concluded that the selection of the 2-piece linear spline model and rejection of the CPH model is flawed by a simple error of not accounting for all three estimated parameters of the spline model[10].

The proposed MON indicates EPA reviewed the TCEQ (2019) assessment but appears to dismiss TCEQ's concern by claiming that the proposed TCEQ assessment had not been peer reviewed.

> TCEQ highlighted uncertainties in the URE arising from what it considered to be errors in the assumptions and calculations used to determine the best model fit of the data. TCEQ's concerns with the EPA's URE derivation have not been peer reviewed and the public comment period closed on September 26, 2019 (Fed Reg 2019 (Tues Dec 17); 84(242):69218).

However, the accounting for all the parameters in a model is not the type of claim that requires peer review to be valid, nor is documenting such a simple error typically considered by peer-reviewed scientific journals a matter worthy of publication and its associated peer-review process. TCEQ clearly demonstrated a simple error by the IRIS assessment due to incorrectly entering into the SAS statistical software the number of parameters estimated for the spline models.[11] TCEQ also provided the corrected AIC and p-values for the spline models that the

[10] The three parameters that were estimated by IRIS (2016) for all 2-piece linear spline models included those representing two slopes and the X-axis value of the "knot" or point of intersection that connects those two slopes

[11] "SAS statistical software used by IRIS required the user to ensure that a correct number of estimated parameters is entered into (or assumed by) that program when fitting a model to data. In the case of 2-piece spline model fits, the SAS program run for the IRIS assessment reflected only two estimated parameters, when in fact three parameters had been numerically optimized for this model (namely, the parameters representing two slopes and the X-axis value of the "knot" or point of intersection that connects those two slopes). This had the effect of making each resulting 2-piece spline model fit appear to be significantly superior to a corresponding simpler log-linear model fit, when in fact both models had statistically equally poor ability to fit the data (TCEQ 2019, pp. 124–129)."

ORD (2019) could have easily verified were correct and used instead of the incorrect IRIS (2016) values. Instead, ORD (2019) relied on the incorrect IRIS (2016) statistics to (in fact, erroneously) judge those fits for the 2-piece spline model to be significantly superior to corresponding fits obtained using the simpler more parsimonious CPH (i.e., "log-linear" risk with cumulative exposures) model.

This basic principle is clearly articulated in the National Research Council report entitled "Models in Environmental Regulatory Decision Making", which states that the strategy to pick the "best model" for regulatory decision making should be "*subject to a penalty function reflecting the number of model parameters, thus effectively forcing a trade-off between improving model fit by adding addition[al estimated] model parameters versus having a parsimonious description*" (NRC, 2007, pp. 174). Importantly, there are no recognized exceptions to the penalty component of the balance incorporated into the AIC metric when applied in a valid procedure for model-selection (Burnham et al. 2002). This general principle is well recognized also to apply specifically to including the estimated "knot" or inflection point from 2-piece linear spline models (Berman et al. 1996, Li et al. 2011; Fearnhead et al. 2019, Gkioulekas et al. 2018, Rodríguez-Domínguez et al. 2018 Molinari et al. 2001). The ORD (2019) memo ignores these well-recognized principles and as a result eliminates the CPH model from consideration in the proposed MON RTR.

We highlight additional important points not previously presented in our comments on the HCl RTR:

- IRIS (2016 Appendix D.3.2. at p D-38) quoted the EPA SAB in justifying the statistical treatment in relation to the knot: "The knot is preselected and is not considered a parameter in these analyses, consistent with SAB's concept of parsimony [footnote 14: "in some setting the principle of parsimony may suggest that the most informative analysis will rely upon fixing some parameters rather than estimating them from the data. The impact of the fixed parameter choices can be evaluated in sensitivity analyses. In the draft assessment, fixing the knot when estimating linear spline model fits from relative risk regressions is one such example" [page 12 of SAB (2015)]".

  However, prior to fitting its two-piece spline model, EPA did not simply "fix" or "select" the position of the knot in that model "rather than estimating" its position, as specified by the SAB. Instead, IRIS tested 20 alternative knots for breast cancer and 70 knots for lymphoid mortality, and then among these, selected knot values that maximized the likelihood of data fit to a corresponding 2-piece spline model. This is not what EPA SAB intended when they suggested that the knot could be "pre-selected" prior to spline-model fitting. EPA SAB

never indicated that IRIS could apply a procedure that violates basic statistical principles, as did the procedure ultimately applied by IRIS.

- In IRIS assessment Appendix D (at p. D-13), Dr. Steenland provided statistics taking into account the knot as a parameter for breast cancer to show this had no substantial effect in that analysis, but a similar examination was not presented in the case of lymphoid cancer. In other words, there was clear acknowledgement and recognition expressed in the IRIS (2016) assessment that each knot value that was used to obtain a final spline-model fit is appropriately interpreted as an estimated parameter. Thus, IRIS should have reported the p-values and AIC taking into account the knot as a parameter for breast and lymphoid cancers in the summary tables of the main report for greater transparency.

- TCEQ (2020a Appendix A) corrected the AIC and p-values reported by IRIS (2016) which are 464.5 and 0.14, respectively, for the IRIS selected 2-piece spline (Table 1, 2, 3). The IRIS (2016) corresponding values for the CPH model are p=0.22 and AIC=464.4. These values are based on IRIS (2016) approaches. Thus, neither the 2-piece spline model nor the CPH model are statistically significant, and the AIC values are similar. Based on statistics alone, the CPH model fits the data similarly to the supralinear 2-piece spline slope, but has the advantage of parsimony (simpler model) and biological plausibility (described below). As described in greater detail in the next section, the CPH model more accurately predicts the observed lymphoid mortalities overall and at lower exposures in the NIOSH study compared to the IRIS (2016) selected 2-piece spline model.

In conclusion, the MON rule is flawed because it depends on the ORD (2019) dose-response analysis, which fails to address and correct the specific, valid statistical considerations described in detail in TCEQ (2020a). When corrected, the CPH model has similar statistical results as the IRIS selected 2-piece spline model (Tables 1, 2, 3). The proposed MON (2019) amendment and the ORD (2019) sensitivity analysis should be corrected to include the CPH model as a relevant considered model. Inclusion of the CPH model results in URE's that are 40-fold lower based on combined cancers and 500-fold lower based on lymphoid mortality (Table 1, 2) based on IRIS calculations and assumptions.

III. ORD (2019) lists SAB (2015) recommendations as the basis for selection of the 2-piece model and alternative models. However, ORD (2019) inclusion of linear regression of categorical data and exclusion of the linear and log-linear CPH models are internally inconsistent and contradictory to SAB (2015) recommendations. The proposed MON rule should include these alternative values, which would result in 15- to 40-fold lower risk values for combined lymphoid and breast cancer incidence, and 140- to 260-fold lower for lymphoid incidence alone. These estimates are based on IRIS (2016) assumptions and methods for deriving upper-bound UREs, and not the EO panel's approach.

ACC agrees with the 2019 MON rule decision to evaluate alternative UREs and central estimates that are consistent with the general principles outlined by SAB (2015). Specifically, ORD (2019) and IRIS (2016) highlight three of SAB recommendations:

- SAB recommended prioritizing functional forms of the exposure that allow regression models with more local fits in the low exposure range (e.g., spline models)
- SAB preferred the use of continuous individual-level exposure data over the use of categorical results.
- SAB advised that any model that is to be considered reasonable for risk assessment must have a dose-response form that is both biologically plausible and consistent with the observed data.

SAB (2015) also indicated that the "principle of parsimony should also be considered".

ORD (2019) Table 1 is misleading, contradictory, and/or unsupported by valid statistical and scientific considerations. Specifically, each of the three Alternative approaches listed involving breast/linear model/individual data (corresponding to the lowest upper bound and central estimates of Unit Risk) is characterized as a "marginal choice", and approaches involving lymphoid/linear model/individual data are not considered at all in Table 1. The choice and characterization of linear-model breast cancer data fits considered in the Memorandum, is explained by ORD (at pp. 4 and 5) as follows:

> A linear model of risk using cumulative EtO dose was examined and provided a statistically significant global fit to the data and a roughly appropriate fit to the categorical data (IRIS, Figure 4-7), however the agreement with the categorical data is poorer in the low-dose region, indicating that the model does not fully meet the SAB goal of providing a local fit to the lower dose data. For the present analysis the linear model is retained as a potentially useful, but **marginally** supported, alternative model. [emphasis added]

> The log-linear (standard Cox) cumulative dose regression model, also provides a statistically significant fit to the global data set but shows notably worse agreement with the plateauing shape of the categorical rates. … Additionally, further data plots for this review indicated that while the log linear model increased roughly linearly over most of the dose range, model predictions, particularly using the upper bound slope estimate, curve sharply upwards at the highest doses – a behavior not indicated by the observed data. Accordingly this model (which would provide a unit risk estimate 13-fold lower than the recommended two-piece spline model) is **not recommended as a reasonable alternative model**. [emphasis added]

whereas the absence of linear fits to individual lymphoid cancer data considered in Table 1 is explained by ORD (2019, at p. 3) as follows:

> Other models fit to the individual level data indicated lower, and sometimes markedly lower, risk estimates but did not provide an appropriate fit to the dose-response pattern in the study data. Among these the log-linear cumulative dose (standard CPH) model and a fully linear model were judged to fit poorly to the data, showing higher AIC values (lower is better), lack of significant fit, and a very inconsistent visual fit to categorical tumor rates (implying minimal increase in risk over the range where the categorical data and other better fitting models indicated substantial risks). Additionally, further evaluation indicates that while the cumulative dose log-linear model showed a shallow linear increase over most of the dose range, model predictions, particularly for the upper bound slope estimate, curve sharply upwards at the highest observed doses. This concave-up behavior is not supported by the observed data.

These rationales for rejecting or marginalizing linear or log-linear Cox Proportional Hazards (PH) model fits to individual-level data contradict SAB recommendations in the following five ways.

1. The fact that ORD (2019) Table 1 includes multiple rows involving alternative models used to fit categorical data, and excludes linear fits to individual-level lymphoid cancer data, contradicts SAB's recommended preference that EPA use continuous individual-level exposure data over categorical results.

2. Linear and standard log-linear CPH fits were rejected or marginalized based on visual evaluations of fits obtained to categorical data, which contradict SAB's recommended preference that EPA use continuous individual-level exposure data over categorical results. These decisions also ignore the caveat appended to the legend of each of the IRIS 2016 Figures 4-2, 4-3, 4-4, 4-5, 4-6, and 4-7 "that, with the exception of the categorical results and the linear regression of the categorical results, the various models have

different implicitly estimated baseline risks; thus, they are not strictly comparable to each other in terms of RR values." This caveat implies the invalidity of all visual evaluations reported in IRIS 2016 comparing plotted categorical data to fitted 2-piece spline models fit to individual-level data vs. linear or standard log-linear CPH model fits to individual-level data. Such comparisons would be valid only if the figures properly accounted for the non-comparability aspect of those model plots along the y-axis noted in the IRIS caveats.

3. All evaluations and comparisons of model-specific statistical fits to individual-level data that involve one of the 2-piece spline models considered in the IRIS assessment are invalid for reasons described in detail in our comments on the EPA's HCl-RTR proposed rule and in section two above. Consequently, conclusions reached in the IRIS assessment that spline model fits to individual-level lymphoid cancer data are superior to corresponding linear or standard log-linear CPH fits are erroneous. None of these models are significant for lymphoid mortality. The IRIS statistical approach to evaluating fits individual-level breast cancer data are similarly erroneous; the consequences of that flaw imply that IRIS erroneously overstated the quality of these fits (all three models, 2-piece linear spline, linear and log-linear) are statistically significant).

4. TCEQ (2020a) uses the models to predict the observed cases for the entire cohort and each of five exposure categories (quintiles[12]). Based on the maximum likelihood estimates (i.e. central estimates), the linear and log-linear CPH models predict not only the overall lymphoid cancers but also the cancers in the lower quintiles of exposure. In other words, the linear and log-linear CPH models are more tuned to local behavior in the data than the IRIS (2016) selected supralinear 2-piece linear spline model. This more objective statistical approach is superior to the subjective comparisons of visual fit based on figures IRIS and SAB (2015) relied on.

5. ORD (2019) rejects the linear and standard log-linear CPH model due to the model behavior at the highest exposure level. In contrast, ORD (2019 at p. 3) admits that fits to categorical risk ratios do "not meet the SAB preference for models fit to the individual data" that excluding highest-dose data to improve fit in the low dose range of regulatory concern is commonly done by EPA in other contexts. Thus, a double standard led ORD (2019) to reject the linear and log-linear CPH models and accept linear regression of the categorical data,

[12] TCEQ (2020a) defines the first quintile as the nine NIOSH workers who died with lymphoid cancer and whose cumulative exposure to EtO (lagged 15 years) was equal to zero. Cumulative exposures to EO lagged 15 years were defined so that quintiles 2 to 5 included the same number of lymphoid cancer deaths (11) in each quintile. This is similar to Steenland et al. (2004) and IRIS (2016) categories of quartiles for lymphoid cancer deaths.

Because the ORD (2019) memo is thus pivotally flawed, so is the proposed MON rule's determination of risk acceptability which relies on the ORD (2019) sensitivity analysis. The ORD (2019) points out (at p. 6) that "It is important to note that this analysis relies entirely on results and equations presented in the final EtO IRIS assessment," and so makes clear that it considered no new information such as that discussed above bearing on critically flawed statistical analysis in the IRIS assessment, or on pertinent new epidemiological and biological data discussed below. Based on statistics alone, the linear and standard log-linear CPH models fit the data just as well as the supralinear 2-piece spline slope for both lymphoid mortality and breast cancer incidence, but have the advantage of parsimony (simpler model) and biological relevance. All three models were not statistically significant for lymphoid incidence, but all three models were statistically significant for breast cancer incidence (Table 3). Thus, there is no statistical basis for excluding the linear and standard log-linear CPH models, and any decision based on poor local fit comparisons with categorical RR (odds ratio) data is flawed. However, the linear and standard log-linear CPH models predict the lymphoid mortalities accurately, whereas the IRIS (2016) selected model over-predicts the observed data statistically significantly. Table 2 and 3 include alternative central estimates and upper bound estimates for the URE for the linear and log-linear CPH models, which should have been included based on statistics and model prediction.

The URE's are based on IRIS (2016) calculations and assumptions, which we do not agree with (see Table 1 footnotes). Two major issues preclude their use as currently derived. First, IRIS (2016) did not demonstrate that the PoD of 1/100 satisfies EPA (2005) cancer guideline definition to be "near the lower end of the observed range." Second, although breast cancer incidence should be considered as part of the weight of evidence for cancer assessment, they should not be used for quantitative cancer risk assessment because of the high potential for bias in the lower exposure range due to underascertainment of cases, most likely among workers who have shorter employment period and are harder to find (Steenland et al. 2003). Although Steenland et al. (2003) considers the subcohort of those interviewed to have complete ascertainment of breast cancer diagnoses, they were selected from and were a subset of the full cohort, which was under-ascertained.

Table 3. Detailed Comparison of IRIS (2016) Risk Values for Models using ORD (2019, Table 1) approach

	Supralinear 2-Piece Linear Spline	Linear regression (forcing to 1 at 0 exposures)	Linear	Standard CPH Linear at EPA POD 1/100
Model of individual data?	Yes	NO	Yes	Yes
IRIS full model name	Linear spline model with knot at 1,600 ppm x days	Linear regression of categorical results	Linear	Log-linear model (standard Cox regression model)
LYMPHOID MORTALITY (M&F)				
IRIS p-value	0.14 corrected from 0.07	0.18	0.13	0.22
EC01 (ppm)	0.0198	0.0607	1.22	1.80
LEC01 (ppm)	5.03 E-03	0.0272	0.318	0.83
Central estimate URE[1] (per ppm)	0.51	0.16	0.01	0.0056
Upper bound URE (per ppm)	1.99	0.37	0.031	0.012
1/1,000,000 RSC ADAF 1.5 (ppt)	0.3	1.8	20.7	53.9
These mortality data are more appropriate basis for quantitative risk assessment than the estimated incidence data below because they are based on the original published data rather than the IRIS (2016) approach, which incorrectly apply the upper bound on the slope for cancer mortality to background incidence rates in a life-table calculation of the excess risk (Sielken and Valdez-Flores, 2009a)				
LYMPHOID INCIDENCE (M&F)				
IRIS p-value	0.14 corrected from 0.07	0.18	0.13	0.22
EC01 (ppm)	7.48 E-03	0.0229	0.462	1.05
LEC01 (ppm)	1.9 E-03	0.0103	0.12	0.49
Central estimate URE (per ppm)	1.34	0.44	0.02	0.0095
Upper bound URE (per ppm)	5.26	0.97	0.0314	0.020
1/1,000,000 RSC ADAF 1.5 (ppt)	0.1	0.7	7.8	31.8

Table 3 (continued)	2-Piece Linear Spline	Linear regression	Linear	Standard CPH
BREAST CANCER INCIDENCE				
IRIS p-value	0.04 corrected from 0.01	0.16	0.01	0.02
EC01 (ppm)	1.38 E-02	0.024	0.054	0.13
LEC01 (ppm)	6.75 E-03	0.011	0.027	0.074
Central estimate URE (per ppm)	0.71	0.42	0.19	0.08
Upper bound URE(per ppm)	1.48	0.91	0.38	0.14
1/1,000,000 RSC not calculated:				
These data are not appropriate for quantitative risk assessment purposes because authors report substantial number of missing cases with higher potential for those with shorter employment missing (Steenland et al. 2003). These data have not been available for independent evaluation by EPA or the public, and, thus, lack transparency, verification and independent analysis. The weight of evidence is weak due to inconsistencies in exposure-response trends and possible biases due to incomplete cancer ascertainment.				
COMBINED LYMPHOID AND BREAST CANCER INCIDENCE				
Central estimate URE (per ppm)	2.1	0.9	0.2	0.1
Upper bound URE (per ppm)	6.1	1.6	0.4	0.15
1/1,000,000 RSC ADAF 1.5 (ppt)	0.1	0.3	1.0	2.9
These data are not appropriate for risk assessment because the breast cancer incidence data are included. In addition, the lymphoid mortality data was transformed to lymphoid incidence using incorrect assumptions (see Table 1 footnotes). EPA provided no justification for the POD of 1/100. TCEQ (2020a) analysis shows that the POD 1/100 for the standard CPH model extrapolates above or in the high range of the experimental data.				

Note: Corrected and IRIS p-values from TCEQ (2020a), IRIS (2016; Tables 4-2, 4-4, 4-6, 4-12, 4-2, 4-13, Appendix D), EC01, LEC01, URE's from IRIS (2016; Tables 4-7, 4-15, 4-17, calculated using Wald-type SE according to IRIS equations in footnote of Table 4-17).

Putting our disagreement with the IRIS (2016) methods aside, the upper bound UREs for combined lymphoid and breast cancer incidence are 15- to 40-fold lower than the IRIS URE when the linear and standard log-linear CPH models are applied. If the UREs are based on lymphoid incidence alone, then the upper bound UREs are 140- and 260-fold lower for the linear and standard log-linear models, respectively. If the UREs are based on lymphoid mortality data – the most appropriate NIOSH cohort data available for quantitative risk assessment – then the upper bound UREs for the alternative linear and log-linear models are 200- to 500-fold lower than for the IRIS (2016) supralinear 2-piece linear spline model (Table 3).

IRIS (2016) acknowledged the lack of mechanistic data to support the biological plausibility of an overall supralinear dose-response, stating "the EPA is not aware of a mechanistic explanation" in response to questions from the USEPA SAB (USEPA 2016 Appendix I, at p. I-29). As will be discussed in detail in the next sections, the log-linear CPH model has the greatest biological plausibility based on the observed data and biological evidence. The linear continuous model has less biological relevance compared to the log-linear CPH model, but is the model that should be considered in place of the linear regression of the categorical data.

In summary, ORD (2019) Table 1 upon which the proposed EPA MON is based is seriously flawed, contradictory to SAB (2015) recommendations, and unsupported by valid statistical and scientific considerations. Consequently, ORD (2019) excludes two models with greater biological plausibility that have comparable statistical and visual fit to the IRIS selected model. These models would have led to 15-40-fold lower upper-bound UREs for combined cancers based on IRIS (2016) methods and assumptions.

IV. The ORD (2019) sensitivity analysis does not consider biological plausibility and consistency based on the results of the epidemiological studies. The overall weak findings suggest a shallow and not a steep exposure response at low exposures, and do not support derivation of one of the highest inhalation URE's.

The ORD (2019) sensitivity analysis focuses solely on statistical and visual fit considerations. In selecting a model for risk assessment, it is important to consider models that are consistent with the observed epidemiological data. It is important to keep into perspective that the relevant epidemiology, despite the large number of human studies published over a forty-year period, indicates that there is only limited evidence of carcinogenicity (IARC 2012; see also ACC comments on the proposed HCl RTR).

While interest has centered on leukemia, other blood related malignancies, and recently breast cancer: (1) there are numerous inconsistencies across the studies, (2) elevated risks above

background are found in isolated studies and the effect size is of small magnitude, and (3) there is an absence of a clear exposure-response relation for any specific cancer type (see ACC comments on the proposed HCl RTR). In a recent systematic literature review and meta-analysis, Marsh et al. (2019) concluded that the most informative epidemiology studies, which were published in the 2000s and 2010s, do not support the conclusion that exposure to EO is associated with an increased risk of lymphohematopoietic cancer or breast cancer. This weight-of-evidence is important to consider in selecting the model because there is no epidemiological evidence that EO is a highly potent human carcinogen.

The epidemiological evidence from the critical NIOSH cohort studies selected by IRIS neither supports the IRIS selection of a supralinear model nor the implication that EO is an extremely potent inhalation carcinogen. Our comments on the proposed HCl RTR rule included a weight of evidence analysis of the lymphohematopoietic cancers that show that the overall evidence for an association with EO is weak and only seen at the highest exposure levels. In the NIOSH cohort, the graphs of the categorical odds ratio data in Section I of these comments reveal that when the confidence intervals are considered, there is no supra-linear spline pattern.

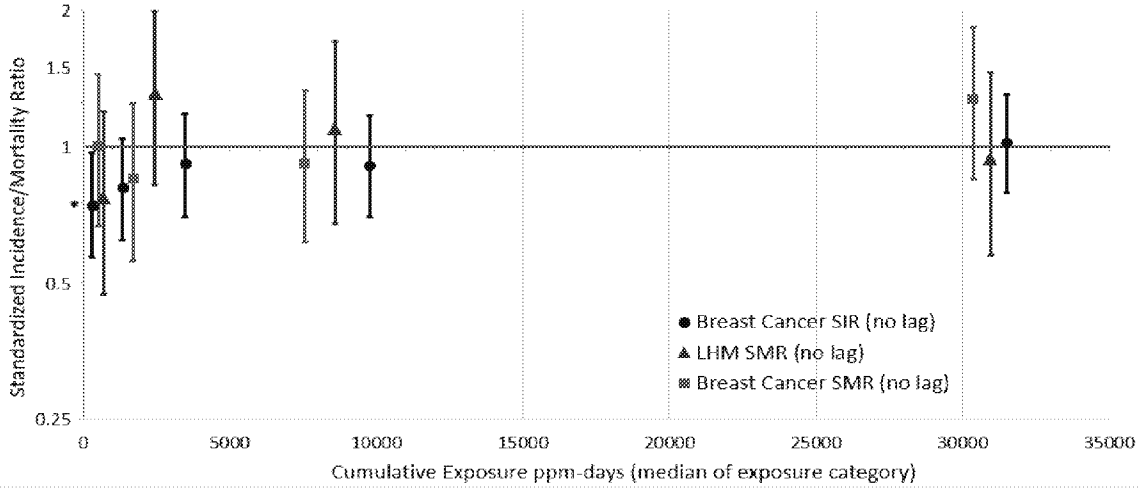
Steenland et al. (2003, 2004) also calculated standardized mortality ratios (SMRs) for lymphohematopoietic cancer mortality[13] and standardized incidence ratios (SIRs) for breast cancer by categories of exposure (i.e. quartiles, quintiles and deciles) for 10-year lag. This type of analysis compares disease incidence or mortality in the exposed population against an external referent group. These SMRs and SIRs estimate excess risk for each category compared to the general population (e.g. life-table analysis for mortality analogous to the life-table analysis IRIS used).[14] There were no statistically significant SIRs or SMRs for breast cancer incidence and lymphohematopoietic cancer mortality (Steenland et al. 2003, 2004). The authors conclude "there was little evidence of cancer excesses" in the mortality data for all cancers examined and no excess of breast cancer in the whole cohort with a non-significant increase in the top quintile of cumulative exposures. These data are not indicative of a highly potent human carcinogen at lower cumulative exposures that the IRIS (2016) URE suggests.

At lower exposures (<647 ppm-days) there is a significant risk deficit (SIRs <1) for breast cancer incidence (Figure 9 black circle). With lag periods included, there are non-significant risk deficits (SIRs and SMRs <1) for both breast cancer incidence below 2,026 ppm days and lymphohematopoietic cancer mortality below 1,199 ppm-days (Figure 10). One possible explanation is that there is a healthy worker effect (HWE) in this cohort. However, the

[13] Lymphohematopoietic cancers include Non-Hodgkin's Lymphoma (NHL), Hodgkin's, myeloma and leukemia. Lymphoid is defined by Steenland et al (2004) to include NHL, myeloma and lymphocytic leukemia. IRIS (2016)
[14] The results of the internal exposure-response analyses in the NIOSH cohort together with an actuarial program (life-table analysis) were used for predicting the extra risks of lymphoid cancer mortality (IRIS, 2016 p 4-9)

epidemiologic literature has shown that HWE is predominately related to populations with shorter follow up and non-cancer causes (Monson 1986; Fox and Collier 1976). Kirkeleit et al. (2013) report no statistically significant healthy worker effect for breast cancer and lymphoid and hematopoietic tissue cancers. Steenland et al. (2004) concluded that "the healthy worker effect would seem an unlikely explanation for the lack of cancer excesses in the exposed versus non-exposed comparisons." These data for breast and lymphoid cancers are not consistent with an extremely potent inhalation human carcinogen, especially at lower exposure levels. In fact, the SIR pattern appears to have a sublinear dose-response at the lower exposures with a statistically significant deficit in the lowest exposure group (Figure 9).

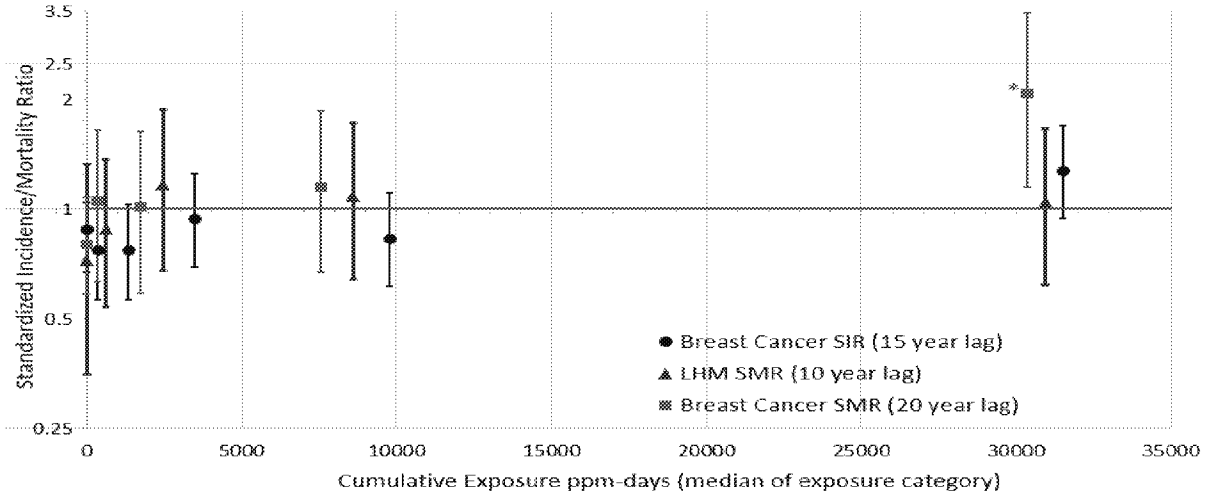
Figure 9 Without a lag period, there was no consistent pattern of increase in SIRs for breast cancer (full cohort) and SMRs for breast cancer and lymphohematopoietic cancers.



Steenland et al (2003, 2004) reported breast cancer SIRs (black circles) and 95% confidence intervals (bars), lymphohematopoietic cancer SMRs (red triangles), and breast cancer SMRs (blue square) by increasing EO ppm-days quintiles (>0 to 647, 647 to 2,026, 2,026 to 4919, 4,919 to 14,620, >14,620 ppm-days), quartiles (>0 to 1,199, 1,200 to 3,679, 3,680 to 13,499, >=13,500 ppm-days), and quartiles (>0 to 646, 647 to 2,779, 2,780 to 12,321, >=12,322 ppm-days), respectively. SIR and 95% confidence interval data come from Table 3 in Steenland et al. 2003, and SMR data comes from Tables 3 and 5 from Steenland et al. 2004. SMR 95% confidence intervals were calculated by first deriving the expected values from the numbers provide in Table 3 (Expected = Observed/SMR) and then using the Mid-P exact test [Miettinen's (1974d) modification, as described in Rothman and Boice (1979).

Figure 10 With lag periods, SIRs for breast cancer and SMRs for breast cancer and lymphohematopoietic cancers do not indicate a supralinear exposure-response at low exposures. The only statistical significance was an increase in breast

cancer SIR at the highest exposure category. Lag periods mean that modeled exposures are more heavily dependent on early historical time predictions which have been shown by Bogen et al. (2019) to underestimate exposures.



Steenland et al. (2003, 2004) reported breast cancer SIRs (black circles) and 95% confidence intervals (bars) with 15 year lag, lympho-hematopoietic cancer SMRs (red triangles) with 10 year lag, and breast cancer SMRs (blue square) with 20 year lag by increasing EO ppm-days quintiles (0 (lagged out), >0 to 647, 647 to 2,026, 2,026 to 4,919, 4,919 to 14,620, >14,620 ppm-days), quartiles (0 (lagged out), >0 to 1,199, 1,200 to 3,679, 3,680 to 13,499, >=13,500 ppm-days), and quartiles (0 (lagged out), >0 to 646, 647 to 2,779, 2,780 to 1,2321, >=12,322 ppm-days), respectively. SIR and 95% confidence interval data comes from Table 3 in Steenland et al. 2003, and SMR data comes from Tables 3 and 5 from Steenland et al. 2004. SMR 95% confidence intervals were calculated by first deriving the expected values from the numbers provide in Table 3 (Expected = Observed/SMR) and then using the *Mid-P exact test* [Miettinen's (1974d) modification, as described in Rothman and Boice (1979).

In conclusion, the IRIS (2016) and ORD (2019) did not check the biological plausibility of the models (i.e. consistency of EO cancer predicted by the models) with the actual epidemiological data. When this is done, the weight of evidence based on the epidemiological data is far more consistent with the CPH rather than a steep 2-piece spline slope.

V. Lymphoid morality in humans is an appropriate health outcome for risk assessment, as is, without transformation to incidence. The overall weak findings of the lymphoid mortality data suggest a shallow and not a steep exposure response at low exposures. These data are not consistent with the IRIS derivation of one of the highest inhalation UREs.

In the NIOSH cohort, there was little evidence of cancer excesses by levels of cumulative exposure for the EO exposed workers versus the general population (Steenland et al. 2003, 2004). A large number of models were considered in their exposure-response analyses and only some sub-analysis showed significance including those using log transformation of cumulative exposure, which IRIS (2016) correctly excluded as biologically implausible. Of the models using cumulative exposures, the strongest trend was seen in male lymphoid mortality. As described in detail in the next section, breast cancer incidence is not an appropriate endpoint based on the weight-of-evidence and quality issues. Therefore, of the critical endpoints selected by IRIS, male lymphoid mortality is the most appropriate endpoint for risk assessment, protective of effects in females who showed no sensitivity.

The cancer risk assessment should be based on lymphoid mortality as the appropriate health outcome without further manipulation to estimate extra risk for lymphoid incidence. The NIOSH study did not collect lymphoid incidence data. IRIS converted lymphoid mortality to lymphoid incidence based on unsupported assumptions that have been shown to introduce error and bias into the analysis (Sielken and Valdez-Flores 2009; Teta et al. 2004). The original data collected by NIOSH should be used without further manipulation that could lead to incorrect characterization of the exposure-response relationship.

It is important to keep into perspective that the relevant epidemiology, despite the large number of studies published over a forty-year period, provide insufficient support based on limited evidence of carcinogenicity (IARC 2012). While interest has centered on leukemia, other blood related malignancies, and recently breast cancer: (1) there are numerous inconsistencies across the studies, (2) elevated risks above background are found in isolated studies and the effect size is of small magnitude, and (3) there is an absence of a clear exposure-response relation for any specific cancer type. In a recent systematic literature review and meta-analysis, Marsh et al. (2019) concluded that the most informative epidemiology studies, which were published in the 2000s and 2010s, do not support the conclusion that exposure to EO is associated with an increased risk of lymphohematopoietic cancer or breast cancer.

In summary, lymphoid mortality without further manipulation to estimate extra risk for lymphoid incidence is an appropriate health outcome for cancer risk assessment. However, the limited weight-of-evidence (IARC 2012, Marsh et al 2019) is important to consider in model selection because epidemiological evidence does not currently support the IRIS derivation of a URE that suggests EO is a highly potent inhalation human carcinogen.

VI. Although useful for consideration of the overall weight-of evidence, breast cancer should not be considered a critical cancer endpoint based on the lack of robust findings in Steenland et al. (2003, 2004) and weight-of-evidence in the epidemiological literature. Breast cancer incidence data should not be used for quantitative risk assessment purposes because of substantial under-ascertainment of cases reported by Steenland et al (2003).

Key reasons why breast cancer should not be considered a critical cancer endpoint for EO are:

1) Neither the NIOSH breast cancer incidence study (Steenland et al. 2003) nor the NIOSH mortality study (Steenland et al. 2004) report an overall excess of breast cancer.
2) The findings are not robust in that they are seen with a certain lag and exposure metric that are not evident with numerous other exposure metrics, models, or lags.
3) The breast cancer incidence findings are at most suggestive, not only due to inconsistencies in the exposure-response, but also due to incomplete cancer ascertainment and the subsequent potential for bias.
4) This disease endpoint is only weakly supported by other epidemiology studies and is inconsistent with others.

The published epidemiology data do not support a supralinear exposure-response relationship for breast cancer. The limited positive findings in the published NIOSH incidence study is seen in the highest exposure category only, not in the lowest or lower levels (Steenland et al. 2003). As described in detail in the previous section, IRIS incorrectly chose a supralinear model based on visual appearance of a limited number of categorical data points without consideration of CIs. This steep exposure-response model is inconsistent with the observed NIOSH data published by Steenland et al. (2003, 2004).

For purposes of hazard assessment and choice of a health endpoint, it is useful to examine all relevant EO studies, even those inadequate for exposure-response analyses. There is no pattern of increase across these six studies and the overall number of observed breast cancers do not exceed expectation, whether based on mortality (113 observed, 116 expected) or incidence

data (372 observed, 425 expected) (see Table 4). In a recent EO meta-analysis based on effect measures from 5 studies, Marsh et al. (2019) also failed to find increased risk for breast cancer among sterilization workers (meta-RR=0.97; 95%CI 0.81-1.18). The most informative cohort by Steenland et al (2003, 2004) and largest contributor reported results very close to expectation (mortality, SMR =0.99) or a significant deficit (incidence SIR=0.87) due to case under ascertainment (Steenland et al. 2004 and 2003, respectively). The findings from EO epidemiology conflict with the IRIS risk values which imply EO is a highly potent carcinogen at lower cumulative exposures.

Table 4 Female Breast Cancer: Overall Observed less than Expected Findings are Not Consistent with IRIS (2016) High Cancer Potency Estimate

Study	Observed	Expected	Obs./Exp.
Coggon et al. 2004	11	13.1	0.84
Steenland et al. 2004	102	103	0.99
Steenland et al. 2003	319	367	0.87*
Mikoczy et al. 2011	41	50.9	0.81
Norman et al. 1995	12	7.00	1.72
Hogstedt et al. 1986	0	---	---
Summary (incident cases only)	372	425	0.88*
Summary (mortality cases only)	113	116	0.97

*Statistically significant but of less interest due to under ascertainment of cases

Although both the NIOSH breast cancer mortality and the breast cancer incidence studies found no increased breast cancer rates overall, they reported some evidence of a trend and increased rates in the highest exposure group for certain forms of exposure modeling but not for others, in the wide variety of statistical analyses conducted (continuous, categorical, cumulative exposure, log of exposure, duration of exposure, lag, no lag, etc.). The authors concluded conservatively, "Our data suggest that ETO is associated with breast cancer…" These suggestive findings were not robust, which would be expected with a potent carcinogen.

The IRIS breast cancer incidence analysis relied on data from the subpopulation of the NIOSH cohort that was interviewed, which required both locating subjects and identifying those diagnosed with breast cancer. Of the 7,576 women in the NIOSH cohort, only 5,139 (68%) were included in the interview portion of the study. The percent non-response was of concern, according to the authors. The majority of these, 22%, could not be located and therefore any breast cancer diagnosis would have been missed. Steenland et al. (2003) indicated that cases lost are more likely to be shorter term (i.e. lower cumulative exposure) employees. Those who work longer (i.e., higher cumulative exposures) stay in the area longer and are more likely to get picked up in the state tumor registries and be found for interview. Shorter duration workers with lower cumulative exposures are more likely to leave the area and not be captured in the overall analyses and less likely to be interviewed.

Steenland et al. (2003) considered the subcohort of women who were interviewed to be "complete". However, there is no way of knowing that the distribution of cases by level of exposure in the subcohort of interviewed breast cancer cases is comparable to the distribution in the fully ascertained total cohort. Steenland indicates that for the "full" cohort, "some women did not have interviews and did not live in states with cancer registries, and it was not possible to estimate the degree of under-ascertainment of the full cohort". Due to the greater difficulty of locating women with short term employment, there is a high potential for bias in missing cases at lower cumulative exposure. Steenland et al. (2003) correctly indicates "there are possible biases due to patterns of non-response and cancer ascertainment which introduce additional uncertainties in the findings," and concluded that the epidemiological evidence was only suggestive for breast cancer.

The question then is whether the subcohort of interviewed population is representative in terms of exposure-response patterns of the fully ascertained cases in the total population. This has been shown not to be the case in some studies (Haneuse, 2016), i.e., participant data alone does not accurately represent the intended study population (participants and non-participants collectively). Kristman et al. (2004) also reported serious bias, even in the case of low loss to follow up, when loss to follow up is not random. Steenland et al. (2003) recognized this possibility and stated they were unable to fully address it. A simple approach was not employed, i.e., to examine in the full population whether the proportion not interviewed was related to level of exposure. **If more cases were missed among those with lower cumulative exposures (shorter term employees), then the data would be biased toward seeing a positive slope and/or elevated risk in the higher exposure groups, as reported by Steenland et al (2003). The lower exposure group(s) would have a deficit of cases**. The possibility of such a bias is strengthened by the finding in this publication of a stronger relationship with duration of employment than with cumulative exposure.

Due to the statistically significant deficit of 0.87 in the overall SIR analysis, Steenland et al. (2003) conducted internal analyses of this exposed population, i.e., workers to workers. Such internal analyses are also conducted when there are concerns about the HWE. However, the epidemiologic literature has shown that HWE is predominately related to shorter follow up and non-cancer causes. (Monson 1986; Fox and Collier 1976). The NIOSH cohort has been followed an average of 25 years. This issue was examined by Gridley et al. (1999) specifically for cancer incidence among Swedish women. The results showed no HWE for breast cancer. Kirkeleit et al. (2013) also report no statistically significant healthy worker effect for breast cancer.

The substantial deficit of cases for breast cancer incidence could have led to non-random cases lost to follow up for the subcohort with interviews. In general, and consistent with Steenland et al. (2003) observations, shorter term workers are more difficult to find for interviews. The high potential for serious bias due to missing cases with lower cumulative exposures compared to higher cumulative exposures renders these data inadequate for quantitative risk assessment purposes. This, together with the unavailability of the breast cancer incidence data to other researchers to independently examine these issues raises quality issues that indicate the data are inappropriate for exposure-response modeling for regulatory cancer risk assessment purposes.

The choice of a supralinear 2-spline model to calculate the URE for breast cancer incidence was heavily weighted by a visual examination of five odds ratios (grouping of data into five exposure categories) from the Steenland et al. (2004) incidence data. The odds ratio in the highest exposure category was significantly elevated (1.87, 95%CI=1.12-3.10). This corresponds to an open-ended cumulative exposure category (greater than 14,620 ppm- days). The remaining odds ratios starting with the lagged-out reference group were 1.00 1.06, 0.99, 1.24, 1.42. Although few in number, the pattern of the odds ratios are clearly not suggestive of supralinearity. The authors observed that when categories of exposure were expanded from five to ten, a different pattern or lack thereof emerged from a decile breakdown (0.88, 1.35, 1.00, 1.00, 1.33, 1.22, 1.40, **1.03**, 1.68, 1.82; bold added to emphasize non-monotonic exposure response).

Using categorical, i.e., grouped data to identify an exposure-response model can be misleading, and the pattern can change as the number of categories are expanded (Valdez-Flores and Sielken, 2013). The odds ratios from the Steenland et al. (2003) breast cancer incidence study appear far from supralinear, as one increases the number of categories examined, as was also illustrated by Valdez-Flores and Sielken (2013). As these authors point out, exposure-response models are best fit with individual rather than summarized data, as recommended by

the SAB and followed by IRIS in their actual modeling. Unfortunately, however, IRIS picked their model *a priori* based on limited categorical data that are not the data modeled.

The exposure-response modeling challenges in the NIOSH publication (and later experienced by IRIS) could be anticipated, given the authors' statement of uncertainty with respect to breast cancer incidence, "The dip in the spline curve in the region of higher exposures suggested an inconsistent or non-monotonic risk with increasing exposure" (Steenland et al. 2003). The other studies that examined breast cancer among women exposed to EO also provide inconsistent results.

Norman et al. examined cancer incidence among 1,132 male and female workers in a medical sterilant plant under active medical surveillance (Norman et al. 1995). The period of potential EO exposure was 1974-1980 and follow up was through 1987. There were 12 breast cancers found among the total of 28 identified cancer cases. Time from first exposure to diagnosis was 11 years or less for each of the 12 cases. These cases would all fall in the NIOSH lagged out group which had a 15-year lag and would therefore be part of the referent group (Steenland et al. 2003). Two of the cases worked at the facility for less than 1 month. Because this was not a well-defined cohort with follow up, the authors used various assumptions and methodologies to calculate person years at risk that yielded a range of SIRs from a statistically significant 2.6 (95% CI: 1.3-5.0) to a non-significant 1.7 (95% CI: 1.0-3.0).

The more recent study by Mikoczy et al. (2011) has been incorrectly cited as supportive of a supralinear association with breast cancer, despite an overall deficit of breast cancer (SIR= 0.81), with or without consideration of a latency period. However, the two higher cumulative exposure groups had statistically significant elevated rates of breast cancer in an *internal* Poisson analysis, due to a substantial and statistically significant deficit of breast cancer in the low dose reference group. This deficit is not explained by the HWE, which is primarily related to non-cancer causes and declines with length of follow up.

This issue was examined by Gridley et al. (1999) specifically for cancer incidence among Swedish women. The results showed no HWE for breast cancer. There are clearly advantages to comparing workers to workers in epidemiology studies to overcome possible biases in external comparisons to the general population. However, there may also be disadvantages to using an internal comparison group that may not be recognized. One danger is selecting a referent group that has an unusual deficit of the disease of interest that creates an artifact of an excess as is illustrated in this study. The IRIS report quantitatively demonstrated the inconsistency of the excesses reported at very low exposures in this population with excesses at only higher exposures in the NIOSH study.

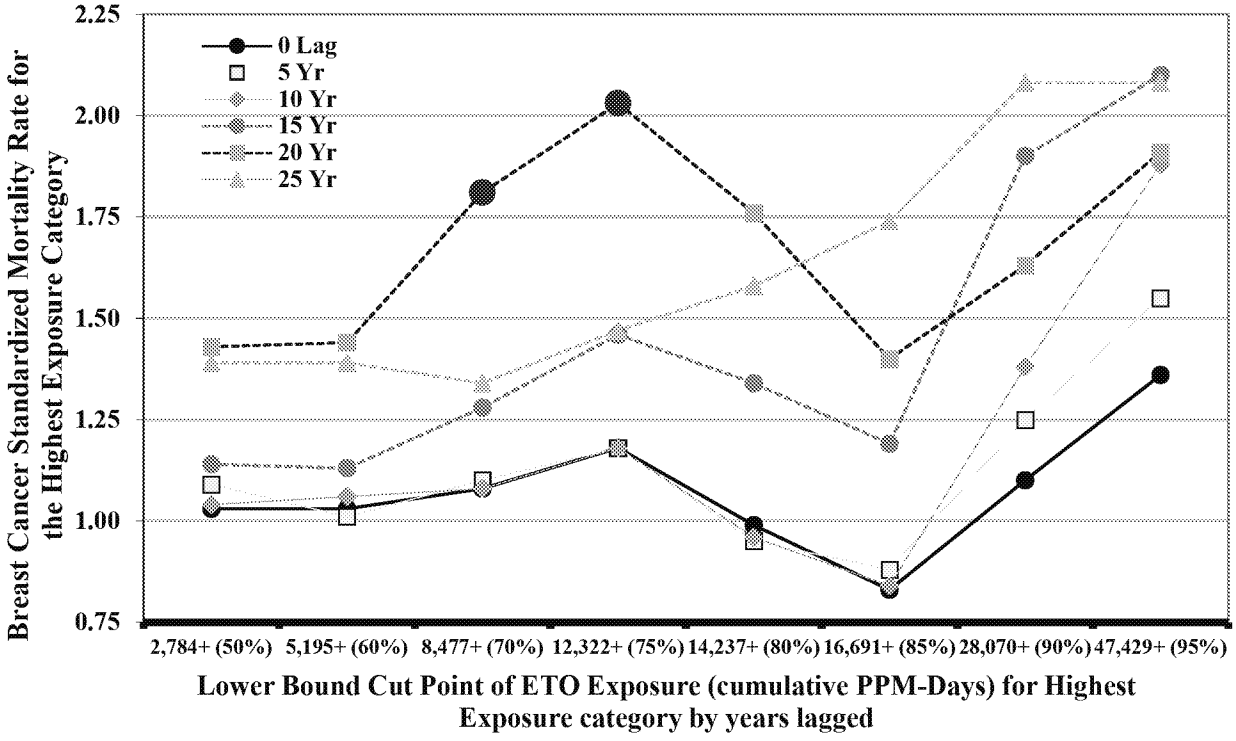
Eight hospitals with EO sterilizer units in England provided 1012 women for a cohort study initially conducted by Gardner et al. (1989) then updated by Coggon et al. (2004). No industrial hygiene data were available before 1977, but exposures were less than 5 ppm after 1977. Peaks of several hundred ppm were known to have occurred from loading and unloading of sterilizers in the hospitals. The authors felt earlier exposures would have been higher and both settings reported peak exposures above the odor threshold (700ppm). Dates of first EO exposure varied from 1962 to 1972 for the hospitals. This study reported no increase in breast cancer (11 deaths observed versus 13.1 expected).

After repeated attempts by the Panel, NIOSH has decided not to share the incidence data from the NIOSH study due to concerns about confidentiality of the workers. This prevents other researchers from evaluating the bias potential of under ascertainment of breast cancer cases, trying alternate or improved methodologies and models, and verifying the NIOSH incidence study and the IRIS exposure-response results, as has been done with lymphoid mortality data.

The authors of the NIOSH study noted that the mean number of ppm-years for the cohort (26.9) is much greater than the median (5.6), indicating a skewed distribution suggesting that there may be a number of subjects with very high cumulative exposures in the highest exposure category. If so, drawing conclusions based on summarized data in the highest exposure category, with the cut-point of 12,322+, could be misleading. Having access to the NIOSH breast cancer mortality data some years ago, we conducted a sensitivity analysis related to this choice of the highest exposure cut point. Our results with mortality data were consistent with observation in the incidence study that breast cancer excess risks showed an "inconsistent or non-monotonic risk with increasing exposure" (Valdez-Flores et al. 2010; Steenland et al. 2003).

In the analyses relied upon by Steenland et al. (2004) in the mortality study (20 yr. lag), cumulative exposure above the 70^{th} and 75^{th} percentiles as the highest exposure category produce statistically significant increases (see large red dots in Figure 11). However, if higher percentile cut-offs were chosen, (i.e., 80%, 85%, 90%, 95%), the SMRs would be lower and none are statistically significant. In other words, the finding of a statistically significant increased risk in the highest exposure category is determined by how the cohort is grouped. Furthermore, it only holds with a 20 yr. lag in which 42 of the 102 breast cancer cases are lagged out from the analysis, leaving only 13 in the highest exposure category. For example, if the highest exposure category was chosen to be 14,237+ ppm-days (the 80^{th} percentile), and the data were analyzed with no lag (black dots), the SMR for the highest exposure category would be 1.0. The consequence of this pattern (or lack thereof) is that it is not possible to identify a definitive cut-off above which excess risk appears.

Figure 11 Increased risk in the highest exposure for breast cancer mortality is determined by how the cohort is grouped and the lag period selected.



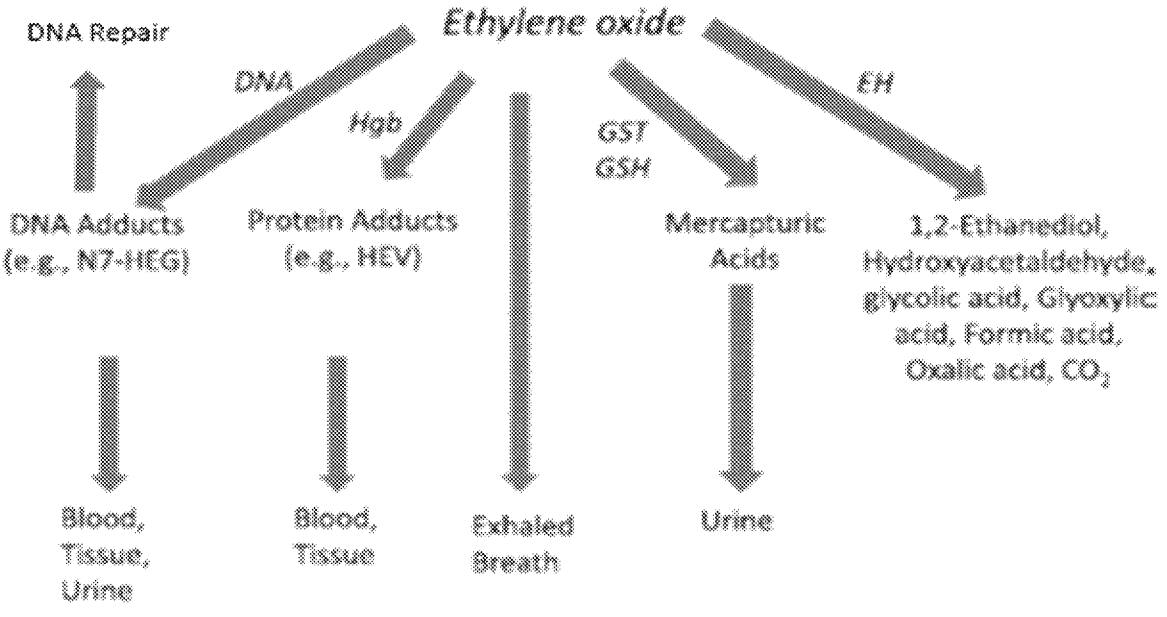
In conclusion, suggestive findings related to breast cancer, with 1) no overall excess, 2) uncertainties in important areas of exposure-response, 3) possible bias associated with case under-ascertainment, and 4) lack of consistency with other studies, should not become the basis for a URE.

VII. IRIS (2016) did not consider the biological plausibility of models based on the biological mode of action and toxicological evidence, which support a shallow linear exposure-response at lower exposures. IRIS has not offered any biologically plausible mode of action analysis accounting for a supralinear dose-response of EO in the low-exposure range. In contrast, considerable experimental mode of action data consistently indicate it is highly implausible that EO operates by supralinear exposure response in the exposure region estimated by IRIS as increasing cancer risks. The IRIS (2016) 1/M risk specific concentration of 0.1 ppt is overly conservative to the point of lacking regulatory utility because it is 4 orders of magnitude lower than average human background (predominately endogenous) exposure levels and variability.

a. The Hypothesized Mode of Action (MoA) and Toxicokinetic Data Indicate a Low-Dose Supralinear Exposure-Response Model Is Biologically Implausible

EO is a direct-acting alkylating agent that forms adducts with hemoglobin, DNA and other cellular macromolecules. As summarized by IRIS (2016), the key molecular initiating event in the default mutagenic MoA for EO has been hypothesized to be EO DNA adduct formation. This can lead to heritable genetic damage cell proliferation, followed by clonal expansion of mutated cells during later stages of cancer development leading to tumor formation.

Figure 12 Disposition and Detoxification of Inhaled EO. DNA= deoxyribonucleic acid; Hgb = hemoglobin; EH = epoxide hydrolase; GSH= glutathione; GST = glutathione-S-transferase; N7-HEG = N7-hydroxyethyl guanine; HEV = N-2-hydroxyethyl valine.



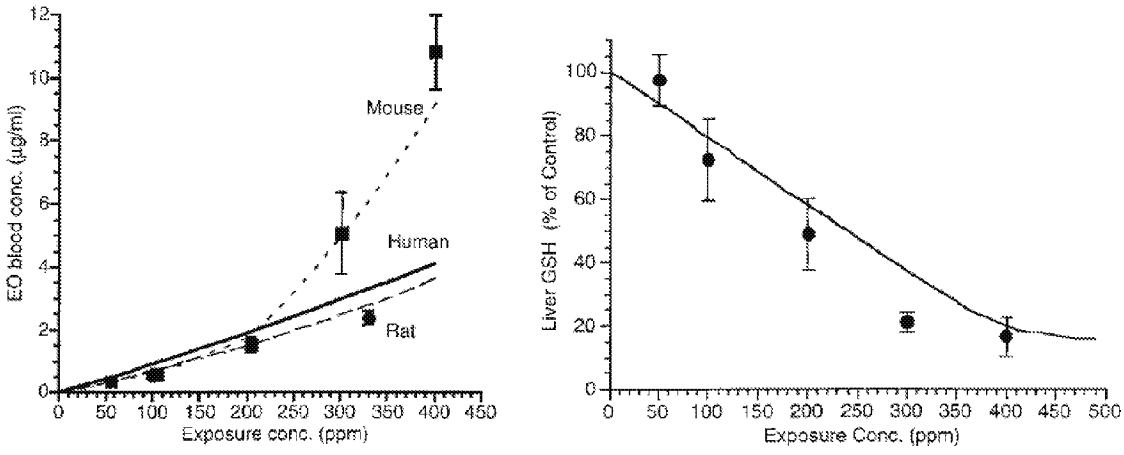
Modified from Kirman & Hays, Reg Toxicol Pharmacol 91: 165-172, 2017

As described previously in ACC comments on the proposed HCl RTR, the metabolic pathways describing the overall disposition of EO are well characterized, i.e., detoxification by direct or enzymatically-mediated conjugation with glutathione and epoxide hydrolase conversion to non-reactive metabolites (Figure 12). In addition, DNA adducts induced by EO are rapidly spontaneously depurinated and/or undergo enzymatic DNA repair. None of these pathways are

expected to operate by a supralinear exposure-response under conditions of low-dose exposure (Filser and Klein, 2018).

In Figure 13, Fennell and Brown (2001) have shown that EO blood concentrations in mice, rats and humans increased linearly with exposures between 50 and 200 ppm. Blood EO increased disproportionately only in mice at exposures exceeding 200 ppm, which was due to substantial depletion of glutathione (GSH) limiting the overall GSH conjugation capacity (GSH, Figure 12).

Figure 13 Toxicokinetics of EO from Fennell and Brown (2001) indicate a linear (not supralinear) exposure response between 50 and >200 ppm



Left panel: Dose-dependent toxicokinetics of ETO in mice, rats and humans.
Right panel: Dose-dependent depletion of liver GSH in mice

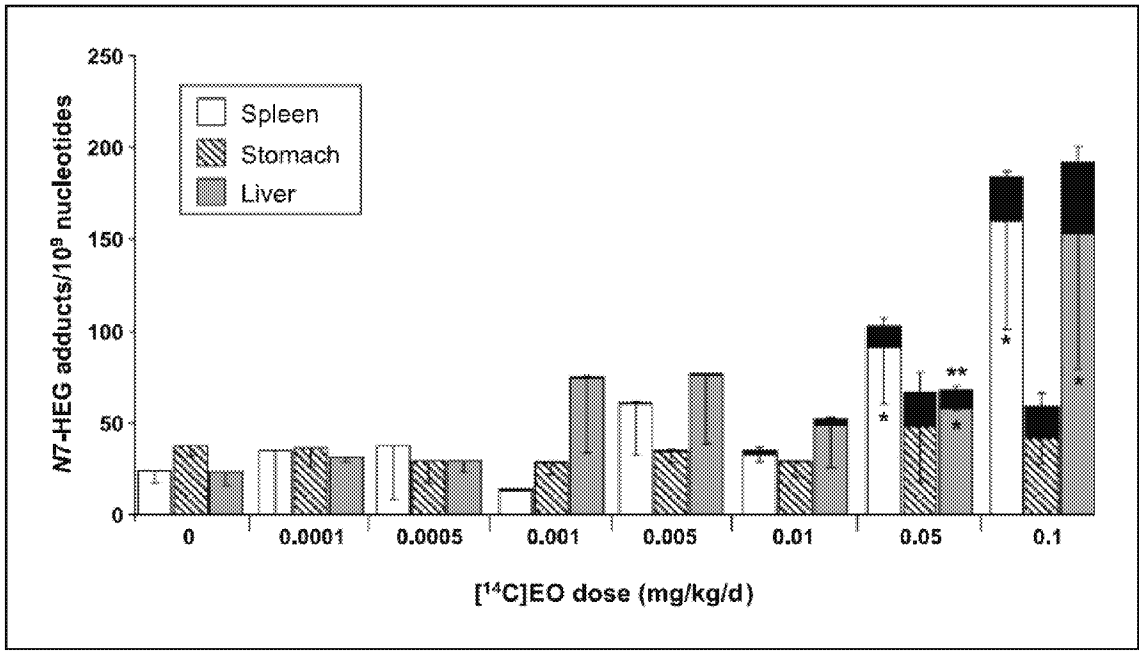
b. Dosimetry of EO DNA adduct formation in rats exposed to EO does not support a low-dose supralinear spline slope

As described above, the key molecular initiating event in the mode of action for EO has been hypothesized to be EO-DNA adduct formation. In comments ACC submitted on the proposed HCl RTR, we highlighted the dose response of formation of the predominate DNA adduct (N7-(2-hydroxyethyl)guanine; N7HEG) formation in rats associated with a wide range of EO exposures (Marsden et al. 2009). We provide additional new perspective based on dose comparisons.

These data indicate that EO did not exhibit any evidence of a supralinear dose-response in the formation of DNA adducts over the range of lower doses evaluated in this study

(Figure 14). Importantly, the total combined endogenous plus exogenous DNA adduct burden was statistically significantly increased only in liver at the second highest 0.05 mg/kg/day dose. Thus, these data indicate that if total DNA adducts were measured by a highly sensitive non-radiolabeled method over a wide range of exogenous EO treatments, such sensitive analyses would not be able to statistically differentiate DNA adducts between controls and treated rats until a dose of 0.05 mg/kg/day. These data are consistent with a sublinear, and more conservatively, a linear exposure response, but not with a supralinear exposure response.

Figure 14 DNA adduct formation, the IRIS (2016) hypothesized key molecular initiating event, do not support a supralinear exposure response. Contribution of endogenously (solid black) and exogenously derived N7-HEG to the total adduct level in tissues of ^{14}C EO-treated rats from Figure 2 of Marsden et al. (2009) is consistent with sublinear, and no more than linear exposure response.



In addition, the lowest dose tested in the rat DNA adduct study, 0.0001 mg/kg/day is approximately 4 orders of magnitude greater than an equivalent EO systemic dose of 4E-08 mg/kg/day in humans inhaling 0.1 ppt EO, the 1/M RSC estimated by the IRIS (2016) risk assessment[15]. These data indicate it is highly unlikely that a 0.1-ppt EO human exposure results in increased tissue DNA adducts until EO exposures far higher than 0.1 ppt are experienced, and

[15] This estimation is based on the following assumptions: 0.1 ppt = 0.18 ng/m^3; humans inhale 20 m^3 air per 24 hr with a 75% respiratory retention (Brugnone et al., 1985), 70 kg body weight.

provide further MoA evidence that EO is not capable of increasing cancer incidence in the low-dose exposure region predicted by the IRIS supralinear spline dose response model.

c. Integration of the tumorigenicity findings from the ethylene and EO rat ethylene and EO carcinogenicity studies also supports a conclusion that the supralinear spline exposure-response model is biologically implausible.

The lack of biological plausibility of the supralinear spline exposure-response is further informed by integration of carcinogenicity findings from the F344 rat carcinogenicity bioassays of ethylene and EO. Because ethylene is metabolized to EO, data from the ethylene bioassay can be considered with that of the rat EO carcinogenicity study (IRIS, 2016 Table 3-5) to further inform the shape of the EO rat carcinogenicity exposure response.

Ethylene has been evaluated for F344 rat (63-80 rats/sex/group) carcinogenicity following a 2-yr inhalation exposure of 300, 1000, and 3,000 ppm (Hamm et al. 1984), and was not carcinogenic at the highest tested exposure. Filser and Klein (2018) used PBPK modeling to estimate that 3,000 or 1000 ppm ethylene exposures in rats (6 hr/day, 5 days/week) were equivalent to 5.52 or 5.26 ppm EO, respectively, in rats. A 40-ppm ethylene exposure was equivalent to 1.26 ppm EO. Importantly, these results were based on modeling of ethylene exposures that resulted in the same levels of N-2-hydroxyethyl valine (HEV) adducts or DNA adducts as those produced by equivalent 6 hr/day, 5 days/week EO exposures. Thus, ethylene was not carcinogenic in F344 rats at a maximum EO equivalent exposure of 5.52 ppm. Overall, the combined ethylene and EO rat carcinogencity data are inconsistent with the hypothesis that EO operates by supralinear exposure-response in the low exposure postulated from the epidemiological exposure-response analysis.

Human background exposure to EO has been estimated by hemoglobin HEV adduct analyses as equivalent to 1,900 ± 1,300 ppt of exogenous EO (Kirman and Hays, 2017). These inhalation-equivalent levels predominately reflect endogenous levels with a relatively small contribution from exogenous background ambient air levels. Recent EO mean background ambient air concentrations measured by EPA approved methods have ranged from 72-154 ppt at various urban, suburban and rural locations in the US with a weighted mean for all values of 113 ppt[16]. Acknowledging this high background endogenous exposure, EPA IRIS (2016) has stated

[16] Recent air concentrations using EPA approved methods were measured to be Denver, CO: 140 (23-580) ppt (Colorado Department of Public Health and Environment, Nov 2018); Chicago, IL: 132 (61-611) ppt (Ramboll, 2019); Chicago, IL (concentration near recently closed sterilization operation): 72 (20-144) ppt (EPA Willowbrook, April, 2019); Grand Rapids, MI: 135 (65-203) ppt (Michigan Department of Environmental Quality, 2019); Atlanta GA (suburban): 111 (17-417) ppt (Georgia Department of Natural Resources, South DeKalb, 2019; Georgia (rural): 104 (22-344) ppt (Georgia Department of Natural Resources, General Coffee Sampling Results, 2019)

that "it is *highly plausible* that the dose-response relationship *over the endogenous range* is sublinear" [emphasis added]. The basis for this conclusion was the knowledge that EO molecular and tissue injury is moderated at low EO exposures by a multiplicity of overlapping biological defenses including primary detoxification by GSH transferase and epoxide hydrolase and secondary intervention of DNA repair (Figure 12).

These data further indicate that it is highly biologically implausible that the contribution of an additional 0.1 ppt exogenous EO to, e.g., an existing 1,900-ppt background endogenous EO exposure, would result in a sudden and biologically unexplained shift to a supralinear dose response and mode of action. This is particularly so considering that such an additional minute exogenous EO exposure is also a very small fraction of even the reasonable variability range of normal human endogenous background EO exposures (1,300 ppt). The coefficient of variation (CV) for humans is 68%. The CVs for rats and mice for endogenous levels measured by Walker et al (1993) in a single study using a single analytical method are 43 and 39%, respectively. The higher CV for humans is consistent with a genetically heterogenous population in humans compared to a genetically homogenous population of rats or mice consuming the same diet. Thus, the variability range of plus or minus 1,300 ppm cannot be attributed solely to experimental variability, and appears to be a reasonable reflection of variability in the human population. Even if one were to make an unrealistic conservative assumption that humans should have the same CV of 40% observed in laboratory rodents, the 0.1 ppt is still a miniscule fraction of a rodent-adjusted variability range of 760 ppt.

In conclusion, as a reactive chemical capable of alkylating DNA, but whose toxicity is modulated by DNA repair and epoxide clearance mechanisms (GSH transferases, epoxide hydrolase) common to rodents and humans, there is no mechanistic rationale to suggest that EO operates by a supralinear exposure response in the low-exposure region projected by IRIS as increasing cancer risks. Indeed, this is also consistent with the observations that there was no excess risk when considering the weight of evidence for the epidemiological studies on EO and the relatively few statistical findings when considering the SIRs, SMRs, odds ratios for the breast cancer incidence and lymphoid mortality data from the Steenland et al. (2003, 2004) studies.

d. The IRIS cancer potency estimate for EO is inconsistent with its weak genotoxic potency.

Several investigations addressed the genotoxicity of EO using *in vitro* and *in vivo* test systems. Given that EO is a direct acting DNA-reactive molecule, it is not surprising that positive results were observed in the majority of these studies. However, EO is a relatively weak mutagen, but the large number of positive studies led to the misunderstanding that it is a potent mutagen (Waters et al. 1999).

A key study indicating supporting the weak mutagenicity of EO comes from the subchronic inhalation study of Manjanatha et al. (2017). These authors investigated the dose-response and temporality for EO-induced mutations at the *cII* locus in the lung tissue of transgenic Big Blue male B6C3F1 mice, a species/strain/sex/tissue where tumors were observed in the EO bioassay. Consistent with mode of action framework analysis objectives, the study design was based on the prediction that if EO-induced mutations were responsible for its tumorigenicity, it should induce mutagenicity in the tumor target tissue in mice at a dose equal to or lower than the tumorigenic dose.

Furthermore, if EO is acting via a mutagenic MoA, the mutant frequency for a neutral gene like *cII* should increase at an early time point, and then continue to increase with continued exposure. Contrary to expectations that are consistent with a mutagenic MoA, no statistically significant increase in mutant frequency or mutational spectrum were observed following 4 weeks of EO exposure (which is considered to be adequate exposure duration for detecting chemically-induced mutations as per OECD test guideline 488), but a significant increase was observed only following 8 or 12 weeks of exposure and only at a concentration (200 ppm) twice the tumorigenic dose in the bioassay. Furthermore, there was no increase in the mutant frequency with exposure duration of 8 and 12 weeks. These data are not consistent with the modified Hill criteria for dose-response and temporality for a mutagenic MoA. This study also demonstrated EO to be weak mutagen with only a small increase in mutant frequency over the background (< 3-fold) even at a concentration twice the tumorigenic dose. Taken together, genotoxicity data support a conclusion that EO is a weak genotoxicant and implausibly associated with a supralinear exposure response at low exposure.

In summary, considerable experimental mode of action data consistently indicate it is highly implausible that EO operates by supralinear exposure response in the exposure region estimated by IRIS as increasing cancer risks. The human data does not support derivation of one of the highest cancer risk values and the animal and biological mode of action strongly suggests a sublinear mode of action over the endogenous range of EO. As a reality check, risk specific concentration of 0.1 ppt is biologically implausible when considering both average human background exposure levels and variability that are 4 orders of magnitude higher.

VIII. The ACC alternative #1 proposal for URE is conservative and has a dose-response form that is both biologically plausible and consistent with the observed data. The rationale for selection of the critical endpoint and point-of-departure are summarized.

Our comments support the conclusion that EPA should not use the EO IRIS Assessment's inhalation RSC of 0.1 ppt to calculate EO risk in its ongoing RTR rulemakings. A more reasonably conservative and scientifically supportable approach to an exposure response analysis is the approach developed by TCEQ (2020a). This alternative approach yields 1/M RSC ranges of 240 – 500 ppt. The point-of-departure and URE values based on the CPH model for lymphoid mortality cases are conservative because extra risk was calculated despite no statistically significant slope in the exposure-response analyses.

The MON includes a range of possible values for cancer risk. We agree with considering a range of values including central estimates, but the ORD (2019) ignored a much more standard statistical model—a CPH model—that has comparable statistical and visual fit to the one selected by IRIS. More importantly, this model has greater biological plausibility fitting EPA SAB's selection criteria for models.

The EO Panel strongly disagrees with ORD's emphasis of visual fit based on a few categorical odds ratio data points which are not the data modeled. ORD ignored the CPH model based on misrepresentations of visual fit and incorrect statistics. We also disagree with including the linear regression of the categorical data which the SAB explicitly stated should not be used because they are not based on the individual data and the full data set should be included.

Our proposed alternative approach is based first on the weight-of-evidence from the epidemiological literature and the Steenland et al. (2003, 2004) papers for determining the critical endpoints, and then we apply the CPH model using the same lag period that IRIS (2016) selected.

As discussed previously, breast cancer incidence is not an appropriate endpoint for risk assessment purposes. The lymphoid mortality was considered by Steenland et al. (2004) to be the more robust finding, with males more sensitive than females. This effect in males, and males and females combined, was modeled using the CPH model with cumulative EO exposure (ppm-days) treated as a continuous variable.

A 1/100,000 extra risk level was estimated consistent with EPA (2005a) cancer risk assessment guidelines on selection of the PoD at the low end of the observable range of responses. When the standard Cox proportional hazard (log-linear) model is used for the NIOSH males-only 15-year lag data, all of the lymphoid mortalities with non-zero exposure occurred **below** the 1 in 100 PoD (Table 5). Therefore, 1 in 100 is not an appropriate PoD for "extrapolation" in the conventional sense.

Table 5 Number of male lymphoid cases from Steenland et al. (2004) with concentrations below the EC(1/100) and EC (1/100,000)

	Male Lymphoid EC 1/100		**Male Lymphoid EC 1/100,000[2]**	
	0-Lag	15-Lag	0-Lag	15-Lag
EC (1/100,000) Env. Conc (ppm)	3.52	5.80	5.83E-03	9.67E-03
Equivalent[1] Occupational Exposure 70 years (ppm-days)	326,105.92	354,399.0[2]	453.4[2]	590.87[2]
Total Number of Deaths	27	27	27	27
Number with zero exposure	0	6	0	6
Number with Non-Zero Exposure below EC	**27**	**21**	**1**	**1**
Percentage of Deaths below EC	100%	**100%**	3.70%	**25.93%**

[1]Equivalent Occupational Exposure 70 years (ppm-days) = EC×(365/240)×(20/10)×365.25×(70-lag)
[2]The maximum occupational exposure concentration for lymphoid deaths was less than 326,106 ppm-days for the unlagged and 137,243 ppm-days for the 15-year lag exposure

A typical POD extrapolates from the edge of the observed range through the unobserved range of the data. Thus, for the NIOSH male only data, it is appropriate to use the model to extrapolate to 1 in 100,000, which is below the 50th percentile of exposure where there is only one lymphoid mortality for subjects with non-zero exposure. IRIS (2016) used a 1% (1 in 100) extra risk for the PoD but did not provide evidence that this level would establish a PoD near the edge of the observed data range. The CPH model has the general form exp(β C) in relation to concentration C and is usually described as a sublinear model. However, it is notable that extra risk predicted by this model is very nearly linear (i.e., proportional to the product β C) at all extra risk levels at or below approximately 0.02. Consequently at all predicted values of excess risk less than about 2% (i.e., about 1 in 50), the CPH model is a very nearly linear model.

Since the 95% lower confidence limit of the 1/100,000 effect concentration (LEC) was considered the most appropriate health protective point-of-departure, we extracted data from TCEQ (2019) that calculated the relevant cancer risk factors based on this same LEC. We then calculated the cancer URE and adjusted the URE by multiplying it by 1.66 to account for the same default age-adjusted default factor (ADAF) to derive a 1/M RSC value of 245 ppt (Tables 6 and 7). The revised TCEQ (2020) rounds this value to 240 ppt.

Table 6 Maximum likelihood estimate (MLE) of the slope parameter, standard error, and deviance (minus two times the log likelihood), and likelihood ratio test statistic corresponding to lymphoid cell line tumors mortality in male workers and both sexes combined

Cancer Outcome	MLE[1]	(SE)[1]	Deviance[1]: -2 × Ln (Likelihood)	Likelihood Ratio Test Statistic[1]	p-value[2] vs. null
Lymphoid Mortality Males only	3.12E-06	(2.61E-06)	356.553	1.052	0.3050
Lymphoid Mortality-Males and Females	2.81E-06	(2.65E-06)	727.899	0.860	0.3537

[1]Values extracted from TCEQ (2019) Tables 7 and 8

[2]p-value calculated based on pre-selecting lag at 15-years based on IRIS (2016) to be more comparable to TCEQ Table 38 correction for IRIS (2016) which also did not consider lag as a parameter

We agree with the proposed MON amendment and the ORD (2019) sensitivity analysis that the central estimate should also be considered in light of additional conservatism IRIS (2016) added by a questionable conversion of the original lymphoid mortality to lymphoid incidence data, and the use of a 15-year lag make the continuous exposure-response models heavily dependent on earlier historical time predictions from the NIOSH exposure regression model, which are likely underestimated (Bogen et al. 2019). When the central estimate of the URE (i.e. maximum likelihood estimate MLE) of 1.03E-03 per ppm is used to derive the URE, then the central estimate for the ADAF-adjusted URE is 1.7E-03 per ppm (Table 7). This central estimate URE value is 1.4 fold lower than the upper bound URE of 2.46 E-03 per ppm. The 1/M RSC value based on the central estimate is 585 ppt.

Table 7 MLE and 95% Lower confidence limit (95%LCL) for Cancer Risk Factors

Cancer Outcome	MLE Environmental Concentration 1/100,000 **ppm**	95% LCL Environmental Concentration (LEC) 1/100,000 **ppm**	MLE URE per **ppm**	95% UCL URE per **ppm**	ADAF adjusted 95% UCL URE (x 1.66) Per **ppb** (per ug/m^3)	ADAF adjusted 1/ M RSC ppt (ug/m^3)
Lymphoid Mortality Males only	9.67E-03	4.07E-03	1.03E-03	2.46E-03	4.1E-06 (2.2E-06)	**245 (0.45)**
Lymphoid Mortality Males & Females	1.32E-02	5.18E-03	7.57E-04	1.93E-03	3.2E-06 (1.8E-06)	312 (0.57)

Additional differences in the ACC recommended approach compared to the IRIS (2016) approach are listed in Table 8, Table 1 footnotes and in previously submitted comments on the proposed HCl RTR rule.

Table 8 Sources of differences between Alternative Approach #1 and EO IRIS Assessment Approach and rationale based on Valdez-Flores et al. (2010)

ACC methods compared to EO IRIS Assessment	Reference and Rationale	Approximate Factor[1]
Extra risk at age 70 instead of 85 years	Valdez-Flores et al. (2010), p. 319 Rationale: IRIS (2016) forces life-table analysis to 85 because of misunderstanding that the cut-off age represents age of life expectancy. Calculating extra risks through age 85 years makes the life table analysis unstable because <1% of cohort lived past 85 involving extrapolation of the fitted models beyond the range of the data upon which they are based. This introduces considerable uncertainty compared to analysis that calculates extra risk through 70 based on cohort information. (Valdez-Flores et al. 2010)	2.3
Extra risk using background rates instead of background mortality rates with lymphoid mortality data (incidence/mortality ratio, $R_{i/m}$).	$R_{i/m}$ = 5.26/1.99 (Valdez-Flores et al. 2010) Rationale: IRIS used the lymphoid mortality model and then applied incorrect assumptions and formulas based on incidence inappropriate for mortality that may significantly alter the exposure-response. The alternative approach relies on original mortality data and assumptions appropriate for mortality (Sielken and Valdez-Flores et al 2009)	2.64

[1]Factor is based on comparison of 0-lag CPH model, so the factor may be slightly different for 15-yr lag CPH model

In summary, our recommended alternative approach #1 results in a 1/M RSC of 240 ppt for the general population including children, and is biologically plausible based on animal data and background levels (predominately endogenous levels) and variability of EtO in humans. This approach is conservative because (a) extra risk was calculated despite no statistically significant slope in the exposure-response analyses; (b) no adjustment was made for likelihood of underestimation of exposures (Bogen et al. 2019[17]); (c) the limited evidence of cancer risk based on the entire body of epidemiologic evidence (Marsh et al. 2019) and in the NIOSH cohort (Steenland et al. 2003, 2004) and (d) the 1/M RSC value of 240 ppt is still substantially below endogenous levels and well within the population variability of endogenous levels.

[17] See ACC comments on the proposed HCl RTR rule

REFERENCES

American Chemistry Council Ethylene Oxide Panel. 2019. Comments on EPA Proposed Amendments to "National Emission Standards for Hazardous Air Pollutants: Hydrochloric Acid Production Residual Risk and Technology Review"

Berman NG, Wong WK, Bhasin S, Ipp E. 1996. Applications of segmented regression models for biomedical studies. Am J Physiol 270(4 Pt 1):E723-32

Bogen KT, Sheehan PJ, Valdez-Flores C, Li AA. 2019. Reevaluation of Historical Exposures to Ethylene Oxide Among U.S. Sterilization Workers in the National Institute of Occupational Safety and Health (NIOSH) Study Cohort. Int J Environ Res Public Health 16(10):1738.

Brugnone F, Perbellini L, Faccini GB, Pasini F, Bartolucci GB, DeRosa E. 1986. Ethylene oxide exposure: Biological monitoring by analysis of alveolar air and blood. Int Arch Occup Environ Health 58:105-112.

Burnham KP, Anderson DR. 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, 2nd ed. Springer-Verlag, New York, NY.

Clemens MR, Remmer H. 1982. Volatile alkanes produced by erythrocytes: an assay for in vitro studies on lipid peroxidation. Blut 45(5):329–335.

Coggon D, Harris EC, Poole J, Palmer KT. 2004. Mortality of workers exposed to ethylene oxide: extended follow up of a British cohort. Occup Environ Med 61(4):358-62.

Colorado Department of Public Health and Environment (CDPHE). 2018. Ethylene Oxide/Terumo BCT Air Sampling Study. Pre- and Post-Control Air Monitoring Report.
Crump KS. 2005. The effect of random error in exposure measurement upon the shape of the exposure response. Dose-Response 3:456-64.

Dutch National Institute for Public Health and the Environment (RIVM). 2019. PROAST. Version 67.0 (menu-driven). https://www.rivm.nl/en/proast.

Einsele H, Clemens MR, Remmer H. 1987. In vitro aging of red blood cells and lipid peroxidation. Arch Toxicol 60(1-3):163–166.

EPA (2005a). U.S. Environmental Protection Agency. Guidelines for Carcinogen Risk Assessment. EPA/630/P-03/001F. U.S. Environmental Protection Agency. Risk Assessment Forum, Washington, DC.

EPA (2005b). U.S. EPA. Supplemental Guidance for Assessing Susceptibility from Early-Life Exposure to Carcinogens. U.S. Environmental Protection Agency, Washington, DC, EPA/630/R-03/003F, 2005

EPA (2008). U.S. Environmental Protection Agency. Reregistration Eligibility Decision for Ethylene Oxide. EPA 738-R-08-003. U.S. Environmental Protection Agency, Office of Pesticide Programs. Washington, DC. March.

EPA (2014) U.S. Environmental Protection Agency (USEPA). National Air Toxics Assessment. https://www.epa.gov/national-air-toxics-assessment/2014-nata-assessment-results.

EPA (2019). U.S. Environmental Protection Agency Outdoor Air Monitoring Data in the Willowbrook Community. February 2019: Ethylene Oxide Concentrations in Outdoor Air - 24-hour averages. https://www.epa.gov/il/outdoor-air-monitoring-data-willowbrook-community
Fearnhead P, Maidstone R, Letchford A. 2019. Detecting changes in slope with an L_0 penalty. J Computat Graph Statist 28(2):265–75.

Fennell TR, Brown CD. 2001. A physiologically based pharmacokinetic model for ethylene oxide in mouse, rat, and human. Toxicol Appl Pharmacol 173: 161-175.

Fetterman, BA; Kim, BS; Margolin, BH; Schildcrout, JS; Smith, MG; Wagner, SM; Zeiger, E. 1997. Predicting rodent carcinogenicity from mutagenic potency measured in the Ames Salmonella assay [Review]. Environ Mol Mutagen 29:312-322.

Filser JG, Kessler W, Artati A, Erbach E, Faller T, Kreuzer PE, Li Q, Lichtmannegger J, Numtip W, Klein D, Pütz C, Semder B, Csanády GA. 2013. Ethylene oxide in blood of ethylene-exposed B6C3F1 mice, Fischer 344 rats, and humans. Toxicol Sci 136(2):344–358.

Filser JG, Klein D. 2018. A physiologically based toxicokinetic model for inhaled ethylene and ETO in mouse, rat, and human. Toxicol Lett 286:54–79.

Fox AJ, Collier PF. 1976. Low mortality rates in industrial cohort studies due to selection for work and survival in the industry. Br J Prev Soc Med 30(4):225-30.

Fu PC, Zic V, Ozimy K. 1979. Studies of ethylene-forming system in rat liver extract. Biochim Biophys Acta 585(3):427–434.

Fukuda H, Ogawa T, Tanase S. 1993. Ethylene production by micro-organisms. Adv Microb Physiol 35:275–306.

Gardner MJ, Coggon D, Pannett B, Harris EC. 1989. Workers exposed to ethylene oxide: a follow up study. Br J Ind Med 46(12):860-5.

Garman RH, Snellings WM, Maronpot RR. 1985. Brain tumors in F344 rats associated with chronic inhalation exposure to ethylene oxide. Neurotoxicology 6: 117-137.

Gkioulekas I, Papageorgiou LG. 2018. Piecewise regression through the Akaike Information Criterion using mathematical programming. IFAC Papers OnLine 51-15:730–35.

Gollapudi BB, Johnson GE, Hernandez LG, Pottenger LH, Dearfield KL, Jeffrey AM, Julien E, Kim JH, Lovell DP, Macgregor JT, Moore MM, van Benthem J, White PA, Zeiger E, Thybaud V. 2013. Quantitative approaches for assessing dose-response relationships in genetic toxicology studies. Environ Mol Mutagen. 54(1):8-18.

Gridley G, Nyren O, Dosemeci M, Moradi T, Adami HO, Carroll L, Zahm SH. 1999. Is there a healthy worker effect for cancer incidence among women in Sweden? Am J Ind Med. 36(1):193-9.

Hamm TE Jr, Guest D, Dent JG. 1984. Chronic toxicity and oncogenicity bioassay of inhaled ethylene in Fischer-344 rats. Fundam Appl Toxicol. 4(3 Pt 1):473–478.

Haneuse S. 2016. Distinguishing Selection Bias and Confounding Bias in Comparative Effectiveness Research. Med Care. 54(4):e23-9.

International Agency for Research on Cancer (IARC). 2012. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. Chemical Agents and Related Occupations. Volume 100F. A Review of Human Carcinogens. World Health Organization. International Agency for Research on Cancer. Lyon, France.

Integrated Risk Information System (IRIS). 2016. Evaluation of the Inhalation Carcinogenicity of Ethylene Oxide (CASRN 75-21-8) In Support of Summary Information on the Integrated Risk Information System (IRIS). U.S. Environmental Protection Agency, Office of Research and Development, National Center for Environmental Assessment. Washington, DC. December.

Johnson GE, Soeteman-Hernández LG, Gollapudi BB, Bodger OG, Dearfield KL,Heflich RH, Hixon JG, Lovell DP, MacGregor JT, Pottenger LH, Thompson CM, Abraham L, Thybaud V, Tanir JY, Zeiger E, van Benthem J, White PA. 2014. Derivation of point of departure (PoD) estimates in genetic toxicology studies and their potential applications in risk assessment. Environ Mol Mutagen. 55(8):609-23.

Kaps M, Lamberson WR. 2004. Biostatistics for Animal Science. Cabi Publishing, Oxfordshire, UK.

Kautiainen A, Törnqvist M, Anderstam B, Vaca CE. 1991. In vivo hemoglobin dosimetry of malonaldehyde and ethene in mice after induction of lipid peroxidation. Effects of membrane lipid fatty acid composition. Carcinogenesis 12(6):1097–1102.

Kirkeleit J, Riise T, Bjørge T, Christiani DC. 2013. The healthy worker effect in cancer incidence studies. Am J Epidemiol 177(11):1218-24.

Kirman CR, Hays SM. 2017. Derivation of endogenous equivalent values to support risk assessment and risk management decisions for an endogenous carcinogen: Ethylene oxide. Regul Toxicol Pharmacol 91:165–172.

Kristman V, Manno M, Côté P. 2004. Loss to follow-up in cohort studies: how much is too much? Eur J Epidemiol 19(8):751-60.

Lawrence GD, Cohen G. 1985. In vivo production of ethylene from 2-keto-4-methylthiobutyrate in mice. Biochem Pharmacol. 34(18):3231–3236.

Li W, He C, Freudenberg J. 2011. A mathematical framework for examining whether a minimum number of chiasmata is required per metacentric chromosome or chromosome arm in human. Genomics 97(3):186-92.

Lynch DW, Lewis TR, Moorman WJ, Burg JR, Groth DH, Khan A, Ackerman LJ, Cockrell BY. 1984. Carcinogenic and toxicologic effects of inhaled ethylene oxide and propylene oxide in F344 rats. Toxicol Appl Pharmacol 76(1):69–84.

MacGregor JT, Frötschl R, White PA, KS, Eastmond DA, Fukushima S, Guérard M, Hayashi M, Soeteman-Hernández LG, Kasamatsu T, Levy DD, Morita T, Müller L, Schoeny R, Schuler MJ, Thybaud V, Johnson GE. 2015a. IWGT report on quantitative approaches to genotoxicity risk assessment I. Methods and metrics for defining exposure-response relationships and points of departure (PoDs). Mutat Res Genet Toxicol Environ Mutagen 783:55-65.

MacGregor JT, Frötschl R, White PA, Crump KS, Eastmond DA, Fukushima S, Guérard M, Hayashi M, Soeteman-Hernández LG, Johnson GE, Kasamatsu T, Levy DD, Morita T, Müller L, Schoeny R, Schuler MJ, Thybaud V. 2015b. IWGT report on quantitative approaches to genotoxicity risk assessment II. Use of point-of-departure (PoD) metrics in defining acceptable exposure limits and assessing human risk. Mutat Res Genet Toxicol Environ Mutagen 783:66-78.

Manjanatha MG, Shelton SD, Chen Y, Parsons BL, Myers MB, McKim KL, Gollapudi BB, Moore NP, Haber LT, Allen B, Moore MM. 2017. Dose and Temporal Evaluation of Ethylene Oxide-Induced Mutagenicity in the Lungs of Male Big Blue Mice Following Inhalation Exposure to Carcinogenic Concentrations. Environ Mol Mutagen 58:122-134.

Mansouri S, Bunch AW. 1989. Bacterial ethylene synthesis from 2-oxo-4-thiobutyric acid and from methionine. J Gen Microbiol. 135(11):2819–2827.

Maronpot RR, Nyska A, Foreman JE, Ramot Y. 2016. The legacy of the F344 rat as a cancer bioassay model (a retrospective summary of three common F344 rat neoplasms). Crit Rev Toxicol 46(8):641-75.

Marsden DA, Jones DJ, Britton RG, Ognibene T, Ubick E, Johnson GE, Farmer PB, Brown K. 2009. Dose-response relationships for N7-(2-hydroxyethyl)guanine induced by low-dose [14C]ethylene oxide: evidence for a novel mechanism of endogenous adduct formation. Cancer Res. 69(7):3052-9.

Marsh GM, Keeton KA, Riordan AS, Best EA, Benson SM. 2019. Ethylene oxide and risk of lympho-hematopoietic cancer and breast cancer: a systematic literature review and meta-analysis. Int Arch Occup Environ Health. 92(7):919-939.

Michigan Department of Environmental Quality (MDEQ). 2019. Interoffice Communication from Amy Robinson to File. Subject: Ethylene Oxide Sampling at Viant. January 4.

Mikoczy Z, Tinnerberg H, Bjork J, Albin M. 2011. Cancer incidence and mortality in Swedish sterilant workers exposed to ethylene oxide: updated cohort study findings 1972-2006. Int J Environ Res Public Health 8(6):2009-19.

Molinari N, Daure J-P, Durand J-F. 2001. Regression splines for threshold selection in survival data analysis. Statist Med 20:237–47.

MON 2019. 40 CFR 463. U.S. Environmental Protection Agency National emission standards for hazardous air pollutants: miscellaneous organic chemical manufacturing residual risk and technology review. Proposed rule. Fed Reg 84(242): 69182-69269.

Monson RR. 1986. Observations on the healthy worker effect. J Occup Med 28(6):425-33.

National Research Council (NRC). 2007. Models in Environmental Regulatory Decision Making. Chapter 5 Model Selection and Use. National Academies Press, Washington DC.

Norman SA, Berlin JA, Soper KA, Middendorf BF, Stolley PD. 1995. Cancer incidence in a group of workers potentially exposed to ethylene oxide. Int J Epidemiol 24(2):276-84.

ORD (2019). U.S. Environmental Protection Agency Office of Research and Development (ORD). 2019. Memorandum from White P, Senior Advisor, to Thayer KA, Director, Office of Research and Development (ORD) Chemical & Pollutant Assessment Division (CPAD), U.S. Environmental Protection Agency, Washington, DC. Subject: Sensitivity of ethylene oxide risk estimates to dose-response model selection 2015 SAB Review. October 18.

Paardekooper LM, van den Bogaart G, Kox M, et al. 2017. Ethylene, an early marker of systemic inflammation in humans. Sci Rep 7(1):6889.

Pottenger LH, Boysen G, Brown K, Cadet J, Fuchs RP, Johnson GE, Swenberg JA. 2019. Understanding the importance of low-molecular weight (ethylene oxide- and propylene oxide-induced) DNA adducts and mutations in risk assessment: Insights from 15 years of research and collaborative discussions. Environ Mol Mutagen 60(2):100-121.

Ramboll Us Corporation (Ramboll). 2019. Ambient Air Monitoring Report. Ethylene Oxide Background Concentrations in Chicago Metro Area. Prepared for Sterigenics US, LLC.
Rodríguez-Domínguez R, Castillo-Vargasmachuca SG, Pérez-González R, Aragón-Noriega EA. 2018. Allometry in *Callinectes bellicosus* (Stimpson, 1859) (*Decapoda: Brachyura: Portunidae*): single-power model versus multi-model approach. J Crutacean Biol 38(5):574-578.
Rothman KJ, Boice JD. 1979. Epidemiologic Analysis with a Programmable Calculator: NIH publication, no. 79-1649. U.S. Department of Health, Education, and Welfare. Public Health Service. National Institutes of Health. Washington, DC.

SAB (2015). U.S. Environmental Protection Agency Science Advisory Board (SAB). 2015. Letter from Thorne PS, Chair, Science Advisory Board, to McCarthy G, Administrator, U.S. Environmental Protection Agency, Washington, DC. Subject: Science Advisory Board Review of the EPA's Evaluation of the Inhalation Carcinogenicity of Ethylene Oxide [Revised External Review Draft - August 2014]. EPA-SAB-15-012. Aug. 7.

Scientific Committee on Occupational Exposure Limits (SCOEL). 2012. Recommendation from the Scientific Committee on Occupational Exposure Limits for Ethylene Oxide. SCOEL/SUM/160. European Commission. Scientific Committee on Occupational Exposure Limits. June.

Sielken RL, Valdez Flores C. 2009a. Life-table calculations of excess risk for incidence versus mortality: ethylene oxide case study. Regul Toxicol Pharmacol 55(1):82-89.

Sielken RL, Valdez Flores C. 2009b. Calculating excess risk with age-dependent adjustment factors and cumulative doses: Ethylene oxide case study. Regul Toxicol Pharmacol 55(1):76–81.

Snellings WM, Weil CS, Maronpot RR. 1984. A two-year inhalation study of the carcinogenic potential of ethylene oxide in Fischer 344 rats. Toxicol Appl Pharmacol. 75(1):105–117.

Soeteman-Hernández LG, Johnson GE, Slob W. 2016. Estimating the carcinogenic potency of chemicals from the in vivo micronucleus test. Mutagenesis 31(3):347–358.

Steenland K, Whelan E, Deddens J, Stayner L, Ward E. 2003. Ethylene oxide and breast cancer incidence in a cohort study of 7576 women (United States). Cancer Causes Control 14(6):531-9.

Steenland K, Stayner L, Deddens J. 2004. Mortality analyses in a cohort of 18 235 ethylene oxide exposed workers: follow up extended from 1987 to 1998. Occup Environ Med 61(1):2-7.

Tates, A.D., Grummt, T., Törnqvist, M., Farmer, P.B., van Dam, F.J., van Mossel, H., Schoemaker HM, Osterman-Golkar S, Uebel C, Tang YS, et al. 1991. Biological and chemical monitoring of occupational exposure to ethylene oxide. Mutat Res 250(1-2):483-97.

Teta MJ, Tran N, Mink PJ, Barraj LM. 2004. Validity of using background leukemia incidence rates with cohort mortality-based potency estimates to calculate excess lifetime risk. Hum Ecol Risk Assess 10:923–938.

TCEQ (2019). Texas Commission on Environmental Quality (TCEQ). Ethylene Oxide Carcinogenic Dose-Response Assessment. CAS Registry Number: 75-21-8. Development Support Document. Proposed, June 28, 2019.

TCEQ (2020a). Texas Commission on Environmental Quality (TCEQ). Ethylene Oxide Carcinogenic Dose-Response Assessment. CAS Registry Number: 75-21-8. Development Support Document. Revised Draft, January 31, 2020.

TCEQ (2020b). Texas Commission on Environmental Quality (TCEQ). Response to public comments received on the 2019 ethylene oxide draft development support document.

Törnqvist M, Gustafsson B, Kautiainen A, Harms-Ringdahl M, Granath F, Ehrenberg L. 1989. Unsaturated lipids and intestinal bacteria as sources of endogenous production of ethene and ethylene oxide. Carcinogenesis 10(1):39–41.

Valdez-Flores C, Sielken RL Jr. 2013. Misinterpretation of categorical rate ratios and inappropriate exposure-response model fitting can lead to biased estimates of risk: ethylene oxide case study. Regul Toxicol Pharmacol 67(2):206-14.

Valdez-Flores C, Sielken RL Jr, Teta MJ. 2010. Quantitative cancer risk assessment based on NIOSH and UCC epidemiological data for workers exposed to ethylene oxide. Regul Toxicol Pharmacol 56(3):312-20.

Vogel, EW; Nivard, MJ. 1998. Genotoxic effects of inhaled ethylene oxide, propylene oxide and butylene oxide on germ cells: Sensitivity of genetic endpoints in relation to dose and repair status. Mutat Res 405: 259-271.

Walker VE, Fennell TR, Upton PB, MacNeela JP, Swenberg JA. Molecular dosimetry of DNA and hemoglobin adducts in mice and rats exposed to ETO. 1993. Environ Health Perspect 99:11–17.

Waters MD, Stack HF, Jackson MA. 1999. Genetic toxicology data in the evaluation of potential human environmental carcinogens. Mutat Res 437:21-49.

Wu KY, Ranasinghe A, Upton PB, Walker VE, Swenberg JA. 1999. Molecular dosimetry of endogenous and ethylene oxide-induced N7-(2-hydroxyethyl) guanine formation in tissues of rodents. Carcinogenesis 20: 1787–1792.

Attachment A

Request for Correction Submitted by ACC to EPA on September 20, 2018

Attachment B

ACC EO Panel Comments on EPA Proposed Amendments to "National Emission Standards for Hazardous Air Pollutants: Hydrochloric Acid Production Residual Risk and Technology Review", April 2019

Attachment C

ACC EO Panel Comments on the TCEQ Proposed Development Support Documents (DSDs) for Ethylene Oxide (EtO) Carcinogenic Dose-Response Assessment, September 2019